

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 19, 2002, 04:57:01 ; Search time 1875 Seconds

(without alignments)  
5937.556 Million cell updates/sec

Title: AI327498  
Perfect score: 532  
Sequence: 1 AATTCGCACTAGACACCC.....CATTAATCCCTGTCATCAT 532

Scoring table: IDENTITY\_NIC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: gb\_ba:\*  
2: gb\_bhg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htggo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Query Match | Score | Length | DB ID | Description |
|------------|-------------|-------|--------|-------|-------------|
|------------|-------------|-------|--------|-------|-------------|

|    |       |      |        |    |            |                       |
|----|-------|------|--------|----|------------|-----------------------|
| 1  | 511.6 | 96.2 | 1412   | 10 | AF079835   | AF079835 Mus muscu    |
| 2  | 511.6 | 96.2 | 1412   | 10 | AF079834   | AF079834 Mus muscu    |
| 3  | 325.2 | 61.1 | 175239 | 2  | AL606903   | AL606903 Mus muscu    |
| 4  | 325.2 | 61.1 | 226957 | 2  | AL606971   | AL606971 Mus muscu    |
| 5  | 101.8 | 19.1 | 1316   | 9  | HDMC46     | HDMC46 Human carbo    |
| 6  | 98.6  | 18.5 | 1344   | 4  | BRCARANVI  | BRCARANVI Human carbo |
| 7  | 68.6  | 12.9 | 107714 | 2  | AL356306   | AL356306 B. taurus mr |
| 8  | 68.6  | 12.9 | 141428 | 9  | AF12841157 | AF12841157 Human DNA  |
| 9  | 67    | 12.6 | 432    | 9  | AF12841157 | AF12841157 Human DNA  |
| 10 | 42.4  | 8.0  | 188036 | 2  | AC091984   | AC091984 Homo sapi    |
| 11 | 42.4  | 8.0  | 195909 | 9  | AF165124   | AF165124 Homo sapi    |
| 12 | 42.4  | 8.0  | 209836 | 9  | AC091926   | AC091926 Homo sapi    |
| 13 | 41.6  | 7.8  | 14568  | 1  | AE004453   | AE004453 Pseudomon    |
| 14 | 41.6  | 7.8  | 1077   | 9  | AF12841156 | AF12841156 Homo sapi  |
| 15 | 40.6  | 7.6  | 82576  | 2  | AC098147   | AC098147 Rattus no    |
| 16 | 40.6  | 7.5  | 28142  | 9  | HSL58B6    | HSL58B6 Human DNA     |
| 17 | 39.4  | 7.4  | 157665 | 10 | AC069018   | AC069018 Mus muscu    |
| 18 | 39.2  | 7.4  | 75314  | 2  | AC023378   | AC023378 Homo sapi    |
| 19 | 39    | 7.3  | 142376 | 2  | AC016959   | AC016959 Homo sapi    |
| 20 | 39    | 7.3  | 178733 | 2  | AF238577   | AF238577 Homo sapi    |
| 21 | 38.8  | 7.3  | 5145   | 2  | AC015315   | AC015315 Drosophila   |
| 22 | 38.8  | 7.3  | 5284   | 3  | DDOKRGEN   | DDOKRGEN Sequence 14  |
| 23 | 38.8  | 7.3  | 182080 | 6  | AC012373   | AC012373 Drosophila   |
| 24 | 38.8  | 7.3  | 303760 | 3  | AE003509   | AE003509 Drosophila   |
| 25 | 38.8  | 7.3  | 164048 | 2  | AC096084   | AC096084 Rattus no    |
| 26 | 38.6  | 7.3  | 91822  | 9  | AL160268   | AL160268 Human DNA    |
| 27 | 38.6  | 7.3  | 134499 | 2  | AC096084   | AC096084 Rattus no    |
| 28 | 38.6  | 7.3  | 164048 | 2  | AC012196   | AC012196 Homo sapi    |
| 29 | 38.4  | 7.2  | 196922 | 2  | AC098158   | AC098158 Rattus no    |
| 30 | 38.2  | 7.2  | 211301 | 2  | AC100752   | AC100752 Mus muscu    |
| 31 | 38    | 7.1  | 2426   | 2  | AC020448   | AC020448 Drosophila   |
| 32 | 38    | 7.1  | 169649 | 9  | AL133328   | AL133328 Human DNA    |
| 33 | 38    | 7.1  | 172904 | 3  | AC007414   | AC007414 Drosophila   |
| 34 | 37.6  | 7.1  | 148415 | 9  | AE003831   | AE003831 Drosophila   |
| 35 | 37.6  | 7.1  | 149572 | 2  | AC016670   | AC016670 Homo sapi    |
| 36 | 37.6  | 7.1  | 149768 | 2  | AC104780   | AC104780 Homo sapi    |
| 37 | 37.6  | 7.1  | 184831 | 2  | AC016239   | AC016239 Homo sapi    |
| 38 | 37.4  | 7.0  | 4969   | 8  | AB017039   | AB017039 Homo sapi    |
| 39 | 37.4  | 7.0  | 7606   | 8  | AB034966   | AB034966 Schistosac   |
| 40 | 37.4  | 7.0  | 7743   | 2  | SPAP5968   | SPAP5968 S. pombe c   |
| 41 | 37.4  | 7.0  | 69774  | 2  | AC105327   | AC105327 Mus muscu    |
| 42 | 37.4  | 7.0  | 2898   | 8  | AF170173   | AF170173 Acetabula    |
| 43 | 37    | 7.0  | 5101   | 1  | PSEMLIAC   | PSEMLIAC Pseudomonas  |
| 44 | 37    | 7.0  | 23630  | 1  | AE008803   | AE008803 Salmonell    |
| 45 | 37    | 7.0  | 23630  | 1  | AE008803   | AE008803 Salmonell    |

## ALIGNMENTS

RESULT 1  
AF079835 1412 bp mRNA linear ROD 04-FEB-1999  
DEFINITION Mus musculus secreted carbonic anhydrase isozyme VI precursor,  
mRNA, complete cds.

ACCESSION AF079835  
VERSION AF079835.1 GI:3421370  
KEYWORDS

SOURCE

ORGANISM

house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS Wang, X.Z., Kuroda, M., Sok, J., Batcharova, N., Kimmel, R., Chung, P.,  
Zinszner, H. and Ron, D.  
TITLE Identification of novel stress-induced genes downstream of chop  
JOURNAL EMBO J. 17 (13), 3619-3630 (1998)

MEDLINE  
98315054  
2 (bases 1 to 1412)

AUTHORS Sok, J., Wang, X.Z., Batcharova, N., Kuroda, M., Harding, H. and Ron, D.  
TITLE CHOP-Dependent stress-inducible expression of a novel form of  
carbonic anhydrase VI  
JOURNAL Mol. Cell Biol. 19 (1), 495-504 (1999)

MEDLINE 9907987  
 REFERENCE 3 (bases 1 to 1412)  
 AUTHORS Sok,J., Wang,X.Z. and Ron,D.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-JUL-1998) Skidball Institute, New York University  
 Medical Center, 540 First Ave., New York, NY 10016, USA

FEATURES  
 source  
 1.1412  
 /organism="Mus musculus"  
 /strain="CD1"  
 /db\_xref="taxon:10090"  
 /tissue\_type="salivary gland"  
 54.125  
 54.1007  
 /note="alternatively spliced"  
 /codon\_start=1  
 /product="secreted carbonic anhydrase isozyme VI precursor"  
 /protein\_id="A012540.1"  
 /db\_xref="GI:3421371"

sig\_peptide  
 CDS

mat\_peptide

BASE COUNT 414 a 343 c 345 g 310 t  
 ORIGIN

Query Match 96.2%; Score 511.6; DB 10; Length 1412;  
 Best Local Similarity 99.2%; Pred. No. 1.9e-140;  
 Matches 514; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

15 AACCCCCCTGCACAGAGATGTCAGTGTGCTTACAGACAAAGTCACTTGTCC 74  
 711 AACCCCCCTGCACAGAGATGTCAGTGTGCTTACAGACAAAGTCACTTGTCC 770  
 75 AAGCCCCGAGTGACCATAGAAACCTGTATGATCAACAACAACACCATTCAA 134  
 771 AAGCCCCGAGTGACCATAGAAACCTGTATGATCAACAACAACACCATTCAA 830  
 135 AATGTTACCCGACACACAGCCCAACACAGGGGTGTGAAGCCAAATTTCTGAAC 194  
 831 AATGTTACCCGACACACAGCCCAACACAGGGGTGTGAAGCCAAATTTCTGAAC 890  
 195 GTCCAGATATGATCTTCTGTACCACTTAACTTAAACATGCAAGAGATTCCTA 254  
 891 GTCCAGATATGATCTTCTGTACCACTTAACTTAAACATGCAAGAGATTCCTA 950  
 255 CAACCTAAGAAACAGAAAGAAAGAAAGAAAGCCGCACTTTGGAGCCGAAATGACAC 314  
 951 CAACCTAAGAAACAGAAAGAAAGAAAGAAAGCCGCACTTTGGAGCCGAAATGACAC 1010  
 315 ACTGCCCCCAGGGCTCACCTGTCTGTAGAAATCTGTGAATCTGGTCTCTCTG 374  
 1011 ACTGCCCCCAGGGCTCACCTGTCTGTAGAAATCTGTGAATCTGGTCTCTCTG 1070  
 375 CTGCTTACAGGAGTCTGTATTAACCCAGAGGAGTCTCTCCGAGAGAAAGTGA 434  
 1071 CTGCTTACAGGAGTCTGTATTAACCCAGAGGAGTCTCTCCGAGAGAAAGTGA 1130  
 435 GCTAGTTAGCCAGAAAGAGATGTGGTGAAGAGGAAAGACTTTCAGCAGAGTG 494  
 1131 GCTAGTTAGCCAGAAAGAGATGTGGTGAAGAGGAAAGACTTTCAGCAGAGTG 1190  
 495 CAAGAAATCAAGCCCTTCATATTCCTGTCATCAT 532  
 1191 CAAGAAATCAAGCCCTTCATATTCCTGTCATCAT 1228

RESULT 2

AF079834 1475 bp mRNA linear ROD 04-FEB-1999  
 LOCUS AF079834  
 DEFINITION Mus musculus stress-inducible intra-cellular carbonic anhydrase  
 isozyme VI mRNA, complete cds.  
 ACCESSION AF079834  
 VERSION AF079834.1 GI:3421368  
 KEYWORDS  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1475)  
 Wang,X.Z., Kuroda,M., Sok,J., Batchvarova,N., Kimmel,R., Chung,P.,  
 Zinsner,H. and Ron,D.  
 Identification of novel stress-induced genes downstream of chop

JOURNAL EMBO J. 17 (13), 3619-3630 (1998)  
 MEDLINE 98315054  
 REFERENCE 2 (bases 1 to 1475)  
 Sok,J., Wang,X.Z., Batchvarova,N., Kuroda,M., Harding,H. and Ron,D.  
 Chop-Dependent stress-inducible expression of a novel form of  
 carbonic anhydrase VI

JOURNAL Mol. Cell Biol. 19 (1), 495-504 (1999)  
 MEDLINE 9907987  
 REFERENCE 3 (bases 1 to 1475)  
 Sok,J., Wang,X.Z. and Ron,D.  
 Direct Submission

JOURNAL Submitted (22-JUL-1998) Skidball Institute, New York University  
 Medical Center, 540 First Ave., New York, NY 10016, USA

FEATURES  
 source  
 1.1475  
 /organism="Mus musculus"  
 /strain="NIH Swiss"  
 /db\_xref="taxon:10090"  
 /cell\_line="NIH-3T3"  
 285.1070  
 /note="alternatively spliced"  
 /codon\_start=1  
 /product="stress-inducible intra-cellular carbonic anhydrase isozyme VI"

BASE COUNT 423 a 370 c 357 g 325 t  
 ORIGIN

CDS

Query Match 96.2%; Score 511.6; DB 10; Length 1475;  
 Best Local Similarity 99.2%; Pred. No. 1.9e-140;  
 Matches 514; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

15 AACCCCCCTGCACAGAGATGTCAGTGTGCTTACAGACAAAGTCACTTGTCC 74  
 774 AACCCCCCTGCACAGAGATGTCAGTGTGCTTACAGACAAAGTCACTTGTCC 833  
 75 AAGCCCCGAGTGACCATAGAAACCTGTATGATCAACAACAACACCATTCAA 134  
 834 AAGCCCCGAGTGACCATAGAAACCTGTATGATCAACAACAACACCATTCAA 893  
 135 AATGTTACCCGACACAGCCCAACACAGGGGTGTGAAGCCAAATTTCTGAAC 194  
 894 AATGTTACCCGACACAGCCCAACACAGGGGTGTGAAGCCAAATTTCTGAAC 953  
 195 GTCCAGATATGATCTTCTGTACCACTTAACTTAAACATGCAAGAGATTCCTA 254  
 954 GTCCAGATATGATCTTCTGTACCACTTAACTTAAACATGCAAGAGATTCCTA 1013  
 255 CAACCTAAGAAACAGAAAGAAAGAAAGCCGCACTTTGGAGCCGAAATGACAC 314  
 1014 CAACCTAAGAAACAGAAAGAAAGAAAGCCGCACTTTGGAGCCGAAATGACAC 1073

```

QY 315 ACTGCCCCAGAGGCTCACCCCTTGCTGTAAGAAATCGTAATGTGATCTCCTTG 374
| | | | |
Db 1074 ACTGCCCCAGAGGCTCACCCCTTGCTGTAAGAAATCGTAATGTGATCTCCTTG 1133
| | | | |
QY 375 CTGCTTCAGAGAGTCTGTGATTAACCCAGAGGAGTTCTCTCCGAGAGAAAGTGGG 434
| | | | |
Db 1134 CTGCTTCAGAGAGTCTGTGATTAACCCAGAGGAGTTCTCTCCGAGAGAAAGTGGG 1193
| | | | |
QY 435 GCTACGTTACCGAAGAAAGAGATGTGGTGAGAGGAAAAAGACTTGACGACGTG 494
| | | | |
Db 1194 GCTACGTTACCGAAGAAAGAGATGTGGTGAGAGGAAAAAGACTTGACGACGTG 1253
| | | | |
QY 495 CAAGAATCAAGCCTTTCATTAATTCCTGTCATCAT 532
| | | | |
Db 1254 CAAGAATCAAGCCTTTCATTAATTCCTGTCATCAT 1291
| | | | |

RESULT 3
AL606903 175239 bp DNA linear HTG 30-NOV-2001
LOCUS Mus musculus chromosome 4 clone RP23-27813, *** SEQUENCING IN
DEFINITION PROGRESS ***, in ordered pieces.
ACCESSION AL606903
VERSION AL606903.7 GI:17381401
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (sites)
REFERENCE 1 (sites)
AUTHORS Garner, P.
TITLE Direct Submission
JOURNAL Submitted (29-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
On Dec 5, 2001 this sequence version replaced gi:16945032.

COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BM27813
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid: L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 175239 bases at least Q40
Consensus quality: 175239 bases at least Q30
Consensus quality: 175239 sum-of-contigs
Insert size: 185643; 1.7% error; agarose-fp
Quality coverage: 14.61x in Q20 bases; sum-of-contigs Quality
coverage: 13.79x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1..175239
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-27813"
/clone_lib="RPCT-23"
/misc_feature
1..175239
/note="assembly-fragment:00292"

BASE COUNT 45250 a 39075 c 40629 g 50285 t

ORIGIN
Query Match 61.1%; Score 325.2; DB 2; Length 175239;
Best Local Similarity 99.1%; Pred. No. 5.6e-85;

```

```

Matches 327; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 203 TATGTACTCTTCTGTACCACTTTACCTAAAAACATGAGAGAGATTTCACAACTAA 262
| | | | |
Db 9470 TCTAGACTCTTCTGTACCACTTTACCTAAAAACATGAGAGAGATTTCACAACTAA 9411
| | | | |
QY 263 GAAACAGAAAGAAACAAAGAAACACCGACCTTTGGAGCCGGGAATATACACTGGGCC 322
| | | | |
Db 9410 GAAACAGAAAGAAACAAAGAAACACCGACCTTTGGAGCCGGGAATATACACTGGGCC 9351
| | | | |
QY 323 CAGGCTCACCCCTGCCCTTGTCTAAGAAATCGTAATGTGAGTCTCTGTCGCTTC 382
| | | | |
Db 9350 CAGGCTCACCCCTGCCCTTGTCTAAGAAATCGTAATGTGAGTCTCTGTCGCTTC 9291
| | | | |
QY 383 ACCGAGTCTCTGATTAACCCAGAGGAGTTCTCTCCGAGAGAAAGTGAAGCTAGCTT 442
| | | | |
Db 9290 ACCGAGTCTCTGATTAACCCAGAGGAGTTCTCTCCGAGAGAAAGTGAAGCTAGCTT 9231
| | | | |
QY 443 ACCCAGAGAGAAAGAGATGTGGTGAGAGGAAAAAGACTTGACGACGTGCAAGAA 502
| | | | |
Db 9230 ACCCAGAGAGAAAGAGATGTGGTGAGAGGAAAAAGACTTGACGACGTGCAAGAA 9171
| | | | |
QY 503 TCAAGCCTTTCATTAATTCCTGTCATCAT 532
| | | | |
Db 9170 TCAAGCCTTTCATTAATTCCTGTCATCAT 9141
| | | | |

RESULT 4
AL606971 226957 bp DNA linear HTG 15-JAN-2002
LOCUS Mus musculus chromosome 4 clone RP23-123120, *** SEQUENCING IN
DEFINITION PROGRESS ***, in unordered pieces.
ACCESSION AL606971
VERSION AL606971.6 GI:18181715
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (sites)
REFERENCE 1 (sites)
AUTHORS Hunt, A.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
On Jan 16, 2002 this sequence version replaced gi:15865162.

COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BM123120
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid: L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 224787 bases at least Q40
Consensus quality: 225129 bases at least Q30
Consensus quality: 225417 bases at least Q20
Insert size: 225857; sum-of-contigs
Insert size: 203349; 5.9% error; agarose-fp
Quality coverage: 8.34x in Q20 bases; sum-of-contigs Quality
coverage: 9.61x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source
1..226957
/organism="Mus musculus"
/db_xref="taxon:10090"

```

```

/chromosome="4"
/clone="RP23-123120"
/clone.lib="RPCT-23"
1..39687
/note="assembly_fragment:05052
fragment_chain:1
clone_end:SP6
vector_side:left"
misc_feature
39988..58240
/note="assembly_fragment:00374
fragment_chain:1"
misc_feature
58341..65122
/note="assembly_fragment:03325
fragment_chain:1"
misc_feature
65223..68735
/note="assembly_fragment:01125
fragment_chain:1"
misc_feature
68836..81936
/note="assembly_fragment:01677
fragment_chain:2"
misc_feature
82037..101775
/note="assembly_fragment:01934
fragment_chain:2"
misc_feature
101876..120162
/note="assembly_fragment:04254
fragment_chain:2"
misc_feature
120263..160511
/note="assembly_fragment:00932
fragment_chain:2"
misc_feature
160612..167303
/note="assembly_fragment:03207
fragment_chain:2"
misc_feature
167404..206353
/note="assembly_fragment:04640
fragment_chain:2"
misc_feature
206454..211015
/note="assembly_fragment:03065
fragment_chain:3"
misc_feature
21116..226957
/note="assembly_fragment:03747
fragment_chain:3"
BASE COUNT 56050 a 35247 c 58829 g 58715 t 1116 others
ORIGIN
Query Match 61.1%; Score 325.2; DB 2; Length 226957;
Best Local Similarity 99.1%; Pred. No. 5.8e-85;
Matches 327; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 203 TATGACTCTTCTGTCACACCTTACCTAAACATGCGAAGAGATTCTACAACCTAA 262
DB 118729 TCTAGACTCTTCTGTCACACCTTACCTAAACATGCGAAGAGATTCTACAACCTAA 118670
QY 263 GAAACGAAAGAAACAAAGAAACCGGCACTTTGGACCGGAAATGACACACCTGGCCC 322
DB 118669 GAAACGAAAGAAACAAAGAAACCGGCACTTTGGACCGGAAATGACACACCTGGCCC 118610
QY 323 CAGGCGTCAACCTGCGCCCTTGTCTAAGGAATCTGGAATGTGGGTCTCTGCTGCGCTTC 382
DB 118609 CAGGCGTCAACCTGCGCCCTTGTCTAAGGAATCTGGAATGTGGGTCTCTGCTGCGCTTC 118550
QY 383 AGCGAGTCTTGATTAACCCAGAGGAGTTCTCTCCGAGAGAAAGTGAAGTCTAGCTT 442
DB 118549 AGCGAGTCTTGATTAACCCAGAGGAGTTCTCTCCGAGAGAAAGTGAAGTCTAGCTT 118490
QY 443 AGCCGAGAGAGAAAGGATGTGGGTGAGAGGAGAAAGACTTGCAGCGAGTGCAGAGAA 502
DB 118489 AGCCGAGAGAGAAAGGATGTGGGTGAGAGGAGAAAGACTTGCAGCGAGTGCAGAGAA 118430
QY 503 TCAAGCCTTTTCAATATTCCTGTCAATCAT 532
DB 118429 TCAAGCCTTTTCAATATTCCTGTCAATCAT 118400

```

```

RESULT 5
LOCUS HUMCA6 1316 bp mRNA linear PRI 31-OCT-1994
DEFINITION Human carbonic anhydrase isozyme VI (CA6) mRNA, complete cds.
ACCESSION M57892.1 GI:179731
VERSION M57892.1 GI:179731
KEYWORDS carbonate dehydratase; carbonic anhydrase isozyme VI; secreted protein.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1316)
AUTHORS Aldred, P., Fu, P., Barrett, G., Penschow, J.D., Wright, R.D.,
Cochlan, J.P. and Bernley, R.T.
Human secreted carbonic anhydrase: cDNA cloning, nucleotide
sequence, and hybridization histochemistry
Biochemistry 30 (2), 569-575 (1991)
91105141
FEATURES
source
Location/Qualifiers
1..1316
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="lp36"
/tissue_type="salivary gland"
<1..1316
/gene="CA6"
/note="G00-125-350; putative"
1..1316
/gene="CA6"
7..933
/gene="CA6"
/EC_number="4.2.1.1"
/note="putative"
/codon_start=1
/product="carbonic anhydrase isozyme VI"
/protein_id="AA051892.1"
/db_xref="GI:179732"
/db_xref="GDB:600-125-350"
/translacion="MRALVLLSLFLGLGQAQOHVSDWTYSGALDEAHWPQHYPAACG
ORSPINLORTVRYNPSLKGIMNGYTOGEEPEPVNNGHTVQIGLPSYRMVAVAG
IYVIAQOHFHNGGASSTSGSEHTVDRIHVEIHIVHSKYKTDIADAPDGLA
VLAARVEKNTPENTYYSNFIISHLANIKYPOQRTLLGLDYOQMLPRYLQHYVYHGS
LTPPCTENVHMFVLADPVKLSRTQVWRLKSLDHRNKTTHNDYRRTQPLKHRYVES
NEPNDEYTLGSEFQYLYKHIEIIDLRLALN"
7..57
/gene="CA6"
/note="G00-125-350; putative"
58..930
/gene="CA6"
/EC_number="4.2.1.1"
/note="G00-125-350; putative"
/product="carbonic anhydrase isozyme VI"
BASE COUNT 362 a 343 c 313 g 298 t
ORIGIN Chromosome 1.
Query Match 19.1%; Score 101.8; DB 9; Length 1316;
Best Local Similarity 67.8%; Pred. No. 4.4e-19;
Matches 158; Conservative 0; Mismatches 72; Indels 3; Gaps 1;
QY 15 ACACCCCCCTGACACAGAGATGTCAGTGGTTGTGCTTAGAGACAAAGTCACTTTGCC 74
DB 667 AGCCTCTCCCTGACACAGAGATGTCAGTGGTTGTGCTTAGAGATTTTGTCAAGCTCTC 726
QY 75 AAGGCCAGGTGTGTGACCATAGAAACTGTATTAGATCACAACAACAACATTCAA 134
DB 727 AGGACACAGSTTTTGGAAAGTGGGAATTCCTTACTGATACACGCAACAAGACATTCAC 786
QY 135 AATGTTATCCGACGACACACGCCAACACACAGGAGTGTGAGACCAATTCTCTGA-- 192
DB 787 AACGATTACCGACGACACGCCAACACAGAGTGTGAGAACACCAACTCCGGAAT 846
QY 193 -ACGTCCAGATATATGACTCTTGTACCACTTACTTAAACAAATGAGAA 244

```

Db 847 CAGGATACACTGCTAGCTGCTGATTCACGTTTACCTACCAATTAAGATTGAGGA 899

## RESULT 6

LOCUS BTRCANRNV1 1344 bp mRNA linear MAM 05-SEP-1996  
DEFINITION B. taurus mRNA for carbonic anhydrase VI.  
ACCESSION X96503  
VERSION X96503.1 GI:1526571  
KEYWORDS CAH6\_bovln gene; carbonic anhydrase VI.  
SOURCE cow.  
ORGANISM Bos taurus

## REFERENCE

AUTHORS Jiang, W., Mollach, J. T. and Gupta, D.

TITLE Sequence of bovine carbonic anhydrase VI: potential recognition sites for N-acetylglucosaminyltransferase

JOURNAL Biochem. J. 318 (Pt 1), 291-296 (1996)

MEDLINE 96358528

AUTHORS Jiang, W.

TITLE Direct Submission  
JOURNAL Submitted (08-MAR-1996) W. Jiang, Pennstate Univ. College of Medicine, Biochemistry & Molecular Biology, P.O. Box 850, Hershey, PA 17033, USA

## FEATURES

source

Location/Qualifiers

1..1344

/organism="Bos taurus"

/db\_xref="taxon:9913"

/tissue\_type="submaxillary gland"

/dev\_stage="adult"

67..108

/gene="CAH6\_bovln"

67..1026

/gene="CAH6\_bovln"

/EC\_number="4.2.1.1"

/codon\_start=1

/product="carbonic anhydrase VI"

/db\_xref="GI:1526572"

/db\_xref="SWISS-PROT:P18915"

/translation="MITLEFLVYGAQAQHEWYSEGVLEDEKHMRLQYPDGGTROSPI

IDLKMKRVNPSIALNLTYGLRQGEPTNNNGHTVOISPSRWMTSPGSOYLA

KQHFHMGDSSEIHSSEHVDGMRYITRIHYHYHSKYSDEKQNEPDLAVLAL

VEYKDYAENWYYSNFISSHEDIRYAGQSTVLDDLDIDMLPDLRLYYSLSLTPPS

CTENWMEYVADIVKLSQTQIEKLENSLHNQNETIÖNNRSTÖPLNHRVVEANFVSH

PHOETVLSKSLHPLYNNIDÖNLEYLRFRITRKKEKYP"

67..1026

/gene="CAH6\_bovln"

1293..1298

/gene="CAH6\_bovln"

polyA\_signal

410 a 325 c 310 g 299 t

BASE COUNT

Query Match 18.5%; Score 98.6; DB 4; Length 1344;  
Best Local Similarity 72.3%; Pred. No. 3.9e-18;  
Matches 128; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Db 15 ACACCCCCCGCAGACAGAAATGCTGAGTGGTTGCTTACAGACAAAGTCACTTGTCC 74  
712 ACTGCTCTCTGAGTACAGAAAACGTCACACTGTTGTGTTAGTAGCAAGTCTCTCT 771  
75 AAGGCCAGGTGCTGACATAGAAACTCGTATGATGATCACAACAACCAACCATTCGA 134  
772 AAGACACAGATGAGAAAGTCTCTTCTTCAATCACCAGACAGACCATTCGA 831  
QY 135 AATGCTTACCGACAGACACCAACCAACAGAGGTGAGTGAAGCCATTTCTCG 191  
832 AATTAATACCGACAGACCATTCGAACCAACAGAGTGTGAAGCCATTCGTCG 888

## RESULT 7

LOCUS AL356306 107714 bp DNA linear HNG 10-JUL-2001  
DEFINITION Homo sapiens chromosome 1 clone RP4-684L19 map p36.21-36.33, \*\*\*  
ACCESSION AL356306  
VERSION AL356306.3 GI:9863760  
KEYWORDS HNG; HTGS\_PHASE1; HTGS\_CANCELLED.  
SOURCE human.  
ORGANISM Homo sapiens

## REFERENCE

AUTHORS McLay, K.

TITLE Direct Submission

JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
On Aug 21, 2000 this sequence version replaced gi:9213829.

## COMMENT

Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquerry@sanger.ac.uk  
Project Information  
Center project name: dj684L19  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: M13; M7815; 2% of reads  
Sequencing vector: plasmid; 108752; 97% of reads  
Chemistry: Dye-terminator ABI; 2% of reads  
Chemistry: Dye-terminator Big Dye; 97% of reads  
Consensus quality: 102398 bases at least Q40  
Consensus quality: 104160 bases at least Q30  
Consensus quality: 105254 bases at least Q20  
Insert size: 106814; sum-of-contents  
Insert size: 125358; 1.4% error; agarose-fp  
Quality coverage: 3.87x in Q20 bases; sum-of-contents quality  
coverage: 3.39x in Q20 bases; agarose-fp

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 20487: contig of 20487 bp in length  
\* 20488 20587: gap of 100 bp  
\* 20588 49082: contig of 28495 bp in length  
\* 49083 49182: gap of 100 bp  
\* 49183 52685: contig of 3503 bp in length  
\* 52686 52785: gap of 100 bp  
\* 52786 60178: contig of 7393 bp in length  
\* 60179 60278: gap of 100 bp  
\* 60279 66201: contig of 5923 bp in length  
\* 66202 66301: gap of 100 bp  
\* 66302 83567: contig of 17266 bp in length  
\* 83568 83667: gap of 100 bp  
\* 83668 96991: contig of 13324 bp in length  
\* 96992 97091: gap of 100 bp  
\* 97092 99503: contig of 2412 bp in length  
\* 99504 99603: gap of 100 bp  
\* 99604 102213: contig of 2610 bp in length  
\* 102214 102313: gap of 100 bp  
\* 102314 107714: contig of 5401 bp in length.

## FEATURES

source

1..107714  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="p36.21-36.33"

```

misc_feature /clone="RP4-684L19"
               /clone.lib="RPC1-4"
               1..20487
               /note="assembly-fragment:00387
               fragment_chain:1"
misc_feature 20588..49082
               /note="assembly-fragment:01016
               fragment_chain:1"
misc_feature 49183..52685
               /note="assembly-fragment:00913
               fragment_chain:1"
misc_feature 52786..60178
               /note="assembly-fragment:00820
               fragment_chain:1"
misc_feature 60279..66201
               /note="assembly-fragment:00242
               fragment_chain:1"
misc_feature 66302..83567
               /note="assembly-fragment:00499
               fragment_chain:2"
misc_feature 83668..96991
               /note="assembly-fragment:00468
               fragment_chain:2"
misc_feature 97092..99503
               /note="assembly-fragment:00315
               fragment_chain:2"
misc_feature 99604..102213
               /note="assembly-fragment:00500
               fragment_chain:2"
misc_feature 102314..107714
               /note="assembly-fragment:00991"
BASE COUNT 27092 a 26474 c 26217 g 27027 t 904 others
ORIGIN

```

```

Query Match 12.9% Score 68.6; DB 2; Length 107714;
Best Local Similarity 74.8%; Pred. No. 5.2e-09;
Matches 86; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

```

```

QY 79 CCCGAGTGGGAGCATGAAAACCTGTATGATCACAACACACACCATTCAAATG 138
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 9586 CCCAGGTTTGACAGTGGAGATTCCTTACGATCCGACACACACCATTCACAGC 95645
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY 139 GTTACCCGACACAGCCCAACACACAGGTTGGGAAGCCATTCTCTGAA 193
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 95646 ATTACCGAGACACCCCTGTAACACACAGAGTGGATCCACTCCGAA 95700
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||

```

```

RESULT 8 AL139415 141428 bp DNA linear PRI 21-JUN-2001
LOCUS AL139415
DEFINITION Human DNA sequence from clone RP3-477M7 on chromosome 1 Contains
            the ENO1 gene for enolase 1, (alpha), the HMGI7 gene for
            high-mobility group (nonhistone chromosomal) protein 17, a zinc
            finger pseudogene, a pseudogene similar to cytochrome c oxidase,
            the C46 gene for carbonic anhydrase VI, ESTs, STSs, GSSs and CpG
            islands, complete sequence.
ACCESSION AL139415
VERSION AL139415.10 GI:13559994
KEYWORDS HMG; C46; carbonic anhydrase; ENO1; enolase; HMGI7; zinc finger.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
            Frankland, J.
TITLE Direct Submission
AUTHORS Submitted (21-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
            requests: clonerequest@sanger.ac.uk
COMMENT On Apr 6, 2001 this sequence version replaced g1:13274269.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the

```

## FEATURES

source

variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chrl>

RP3-477M7 is from the library RPC1-3 constructed by the group of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pCYPAC2

This sequence is the entire insert of clone RP3-477M7.

## Location/Qualifiers

```

1..141428
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP3-477M7"
/clone.lib="RPC1-3"
293..435
/clone.lib="RPC1-3"
436..747
/note="AluYo repeat: matches 1..140 of consensus"
748..894
/note="AluYo repeat: matches 1..310 of consensus"
994..1052
/note="LINE3 repeat: matches 5987..6047 of consensus"
1089..1264
/note="THEIC repeat: matches 1..133 of consensus"
1265..1577
/note="AluX repeat: matches 2..311 of consensus"
1578..1800
/note="THEIC repeat: matches 133..371 of consensus"
1941..2213
/note="AluX repeat: matches 21..293 of consensus"
2353..2454
/note="MIR repeat: matches 80..172 of consensus"
complement(2444..2843)
/note="match: GSS: Em:AQ076041"
complement(2455..2645)
/note="match: GSS: Em:AQ075536"
complement(2469..2684)
/note="match: GSS: Em:AQ065754"
complement(2586..2904)
/note="match: GSS: Em:AQ05252"
2794..2958
/note="MERB repeat: matches 1..178 of consensus"
2979..3115
/note="FLAM_A repeat: matches 1..125 of consensus"
3141..3430
/note="AluX repeat: matches 1..290 of consensus"
4013..4159
/note="HERVL repeat: matches 5615..5764 of consensus"
4162..4320
/note="MUT2 repeat: matches 13..188 of consensus"
4321..4614
/note="AluX repeat: matches 1..296 of consensus"
4615..4706
/note="MUT2 repeat: matches 188..275 of consensus"
4790..4944
/note="AluY repeat: matches 130..286 of consensus"

```

CDS

```
/note="match: STS: Em:G05574"
complement(join(6779..6848,8306..8364,8654..8762,
9310..9511,10704..10901,11698..11920,12536..12669,
13407..13476,15871..15929,17710..17405,20243..20327))
/genre="EN01"
/note="match: proteins: Sw:P51913 Tr:Q9W7L0 Tr:Q9W7L2
Tr:Q9XSJ4 Sw:P06733 Sw:P17182 Sw:PI9140 Sw:P04764
Sw:O57391 Sw:P08734"
/evidence=not_experimental
/product="d1477W7.1 (enolase 1, (alpha))"
/protein_id="CAC42425.1"
/db_xref="GI:14530765"
/translation="MSILKIHAREIFDSRGNPTVEVDLFTSKGLFRAAPSGASTGI
EALFELNDNDKTRMGKGVSKAVEHINKTAPALVSKLNTGEOKIDKLIEMDGTG
KSRFGANAILIGVSLAVCKAGAEVKEVPLTIRHIALDAGNSFVLLPVPAPVINGSHAI
NKLAOMEFMILIPVGAANPREAMRIGAEVYHIALKVITREYKGRATVNGDEGTPAPHI
ENEGKEELIKTAIGKAGRTDKYVIGMDYAASEFRSGKYIDLPKSPDDPSRTISPDO
ADIKSEFIDKDFVAVSTLEPDDDDGAAQKFTTASAGIQVVGDDLYTNPKRLKAVN
KSCNCLIKLVNQIGSTVSLQACKLAQANGGVAVSVRSGETEDTEIADLYVGLCTAGN
IKTGACRSERLEKATYNOLRIIEELGSKAKFAGRNFRNPLAK"
7323..7602
/note="AlusX repeat: matches 31..310 of consensus"
7605..7700
/note="AluUo/PRAM repeat: matches 209..304 of consensus"
8306..8762
/note="match: GSS: Em:A452511"
9055..9140
/note="PRAM C repeat: matches 21..106 of consensus"
9747..10024

Query Match 12.9%; Score 68.6; DB 9; Length 141428;
Best Local Similarity 74.8%; Pred. No. 5.4e-09;
Matches 86; Conservative 0; Mismatches 29; Indels 0; Gaps 0

QY 79 CCCAGGTGTGCACCATGAAACCTGTTATGATGACACACAACAACACCATTCGAATG 138
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116281 CCCAGGTTTGAAAGCTGGAGAAATTCCTTACTGATGATCACCACACAGACATCCACACAGC 116340

QY 139 GTTACCCGACACACAGCCCAACACACACAGGAGGTGTGGAAGCCAAATTTCTGGA 193
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116341 ATTACCCGACGACCCAGCCCTGTGAACACACAGAGTGTGGAATCCAACTCCCGGA 116395

RESULT 9
AF12841157 432 bp DNA linear PRI 02-JAN-2000
LOCUS AF12841157
DEFINITION Homo sapiens carbonic anhydrase VI gene, exon 7.
ACCESSION AF1284117
VERSION AF128417.1 GI:6652973
KEYWORDS
SEGMENTS
SOURCE human.
ORIGIN
7 of 8
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 432)
Grubb,D.J.
Direct Submission
Submitted (16-FEB-1999) Howard Florey Institute, University of
Melbourne, Royal Parade, Melbourne, Vic 3133, Australia
FEATURES
source
1..432
/organism="Homo sapiens"
/db_xref="taxon:9606"
284..397
/number=7
BASE COUNT 102 a 133 c 97 g 100 t
ORIGIN
Query Match 12.6%; Score 67; DB 9; Length 432;
Best Local Similarity 73.9%; Pred. No. 7.7e-09;
Matches 85; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
```



```

OY 79 CCCAGGTGTGACCATAGAAAACCTGTATGATCACAACACACCATTCCTCAATG 138
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 279 CCCAGGTGTGACCATAGAAAACCTGTATGATCACAACACACCATTCCTCAATG 338
OY 139 GTTACCGCAGACACACACACACACACAGGCGTGTGAAGCAATTCCTGAA 193
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 339 ATTACCGCAGACACACACACACACAGGCGTGTGAAGCAATTCCTGAA 393

RESULT 10
AC091984/ 188036 bp DNA 1linear HTG 09-JUN-2001
LOCUS Homo sapiens chromosome 5 clone RP11-569B13, WORKING DRAFT
DEFINITION
SEQUENCE: 30 unordered pieces.
AC091984
AC091984.1 GI:14333920
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
2 (bases 1 to 188036)
DOE Joint Genome Institute.
Direct Submission
Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web Site: http://www.jgi.doe.gov

Project Information
Center Project Name: 624898
Center clone name: RPCI-11_569B13

Summary Statistics
Consensus quality: 149093 bases at least Q40
Consensus quality: 171742 bases at least Q30
Consensus quality: 175083 bases at least Q20
Estimated insert size: 174490; agarose-fp estimation
Estimated insert size: 185136; sum-of-contigs estimation
Quality coverage: 6.42 in Q20 bases; agarose-fp estimation
Quality coverage: 6.05 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1488: contig of 1488 bp in length
* 1489 1588: gap of unknown length
* 1589 2840: contig of 1252 bp in length
* 2841 2940: gap of unknown length
* 2941 4281: contig of 1341 bp in length
* 4282 4382: gap of unknown length
* 4382 5562: contig of 1181 bp in length
* 5562 5663: gap of unknown length
* 5663 6777: contig of 1115 bp in length
* 6777 6878: gap of unknown length
* 6878 7936: contig of 1059 bp in length
* 7937 8036: gap of unknown length
* 8037 9106: contig of 1070 bp in length
* 9107 9207: gap of unknown length
* 9207 10484: contig of 1277 bp in length
* 10484 10583: gap of unknown length
* 10583 12565: contig of 1982 bp in length

```

```

12665: gap of unknown length
* 12666 15417: contig of 2752 bp in length
* 15418 15517: gap of unknown length
* 15518 20425: contig of 4908 bp in length
* 20426 20525: gap of unknown length
* 20526 25068: contig of 4543 bp in length
* 25069 25168: gap of unknown length
* 25169 30263: contig of 5095 bp in length
* 30264 30363: gap of unknown length
* 30364 34974: contig of 4611 bp in length
* 34975 35074: gap of unknown length
* 35075 43270: contig of 8196 bp in length
* 43271 43370: gap of unknown length
* 43371 49361: contig of 5991 bp in length
* 49362 49461: gap of unknown length
* 49462 56288: contig of 6827 bp in length
* 56289 56388: gap of unknown length
* 56389 63723: contig of 7335 bp in length
* 63724 63823: gap of unknown length
* 63824 69752: contig of 5929 bp in length
* 69753 69852: gap of unknown length
* 69853 78189: contig of 8337 bp in length
* 78190 78289: gap of unknown length
* 78290 85925: contig of 7636 bp in length
* 85926 86025: gap of unknown length
* 86026 92953: contig of 6528 bp in length
* 92954 93053: gap of unknown length
* 93054 101358: contig of 8305 bp in length
* 101359 101458: gap of unknown length
* 101459 108969: contig of 7511 bp in length
* 108970 109069: gap of unknown length
* 109070 117359: contig of 8280 bp in length
* 117360 117459: gap of unknown length
* 117460 126188: contig of 8729 bp in length
* 126189 126288: gap of unknown length
* 126289 138034: contig of 11746 bp in length
* 138035 138134: gap of unknown length
* 138135 150053: contig of 11919 bp in length
* 150054 150153: gap of unknown length
* 150154 169677: contig of 19424 bp in length
* 169678 169679: gap of unknown length
* 169679 188036: contig of 18359 bp in length.
Location/Qualifiers
1.188036
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-569B13"
/clone_1b="RPCI human BAC library 11"
BASE COUNT 57018 a 36648 c 33629 g 57817 t 2924 others
ORIGIN
Query Match 8.0%; Score 42.4; DB 2; Length 188036;
Best Local Similarity 66.3%; Pred. No. 0.32;
Matches 61; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
OY 219 CACCTTACTAATAAAGACGAGAGATCTCAACCTAAGAAAGAGAAACA 278
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 61874 CACCTGTATTGAATAAAGAACTAGACATCTTCAATTGAGGCAAGAAACA 61815
OY 279 AAGAAGACCGGACATTTGGAGCCGGAATG 310
    ||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||
DB 61814 AATGGAACTAGCATTTTGGAAAGAGAACTG 61783

RESULT 11
AF165124 195909 bp DNA 1linear PRI 07-JUN-2001
LOCUS Homo sapiens chromosome 5q31.1-q33.1 clone BAC djn082c10 containing
DEFINITION GABRG2 gene, complete sequence.
ACCESSION AF165124
VERSION AF165124.1 GI:5738137
HTG.

```



| ORGANISM              | Homo sapiens   |
|-----------------------|--|
| REFERENCE             | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.          |
| REFERENCE             | 1 (bases 1 to 209836)  |
| REFERENCE             | DOE Joint Genome Institute and Stanford Human Genome Center.   |
| REFERENCE             | 2 (bases 1 to 209836)  |
| REFERENCE             | DOE Joint Genome Institute.  |
| REFERENCE             | Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA |
| REFERENCE             | 3 (bases 1 to 209836)  |
| REFERENCE             | DOE Joint Genome Institute and Stanford Human Genome Center.   |
| REFERENCE             | Submitted (30-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA                                 |
| REFERENCE             | 4 (bases 1 to 209836)  |
| REFERENCE             | DOE Joint Genome Institute and Stanford Human Genome Center.   |
| REFERENCE             | Submitted (21-SEP-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA                                 |
| REFERENCE             | On Sep 21, 2001 this sequence version replaced gi:15375175.  |
| REFERENCE             | Draft Sequence Produced by DOE Joint Genome Institute  |
| REFERENCE             | www.jgi.doe.gov  |
| REFERENCE             | Finishing Completed at Stanford Human Genome Center  |
| REFERENCE             | www.shgc.stanford.edu  |
| REFERENCE             | Quality: Phrap Quality >=40 99.8% of Sequence;   |
| REFERENCE             | Estimated total Number of Errors is 0.3.   |
| REFERENCE             | SMS Content:   |
| REFERENCE             | SHGC-79397 G52187.   |
| FEATURES              | Location/Qualifiers  |
| FEATURES              | 1..209836  |
| FEATURES              | /organism="Homo sapiens"   |
| FEATURES              | /db_xref="taxon:9606"  |
| FEATURES              | /chromosome="5"  |
| FEATURES              | /clone="RP11-28066"  |
| BASE COUNT            | 66857 a 39092 c 37729 g 66158 t  |
| ORIGIN                |  |
| Query Match           | 8.0%; Score 42.4; DB 9; Length 209836;   |
| Best Local Similarity | 66.3%; Pred. No. 0.33; 31; Indels 0; Gaps 0;   |
| Matches               | 61; Conservative 0; Mismatches 31; Indels 0; Gaps 0;   |
| QY 219                | CACCTTACTTAAAAACATGCGAGAGAGATCTTACACCTTAAAGAAAGAGAAACA 278   |
| Db 65753              | CACCTGTATTATGAAATTAATGAACTACGATCTTCTCAATTGAGCGCAAGAGAAACA 65812  |
| QY 279                | AAGAAGACCGCAGCTTTGGAGCGGGAAG 310   |
| Db 65813              | AATGGGAGCTACGATTTTGGACAGGAGCT 65844  |
| RESULT 13             |  |
| AE004453              |  |
| LOCUS                 | 14568 bp DNA linear BCT 30-AUG-2000  |
| DEFINITION            | Pseudomonas aeruginosa PA01, section 14 of 529 of the complete genome.   |
| ACCESSION             | AE004453 AE004091  |
| VERSION               | AE004453.1 GI:9945978  |
| KEYWORDS              | .  |
| SOURCE                | Pseudomonas aeruginosa.  |
| ORGANISM              | Pseudomonas aeruginosa   |
| REFERENCE             | Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.  |
| REFERENCE             | 1 (bases 1 to 14568)   |
| REFERENCE             | Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warren,P.,   |
| REFERENCE             | Hickey,M.J., Brinkman,F.S., Huftnagle,W.O., Kowalik,D.J., Lagrou,M.,   |
| REFERENCE             | Gardner,R.L., Goltzy,R., Tolentino,E., Westbrook-Adman,S., Yuan,Y.,  |
| REFERENCE             | Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Lardig,K., Lim,R.,  |
| REFERENCE             | Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.  |
| REFERENCE             | Complete genome sequence of Pseudomonas aeruginosa PA01, an  |

JOURNAL Nature 406 (6799), 959-964 (2000)

MEDLINE 20437337

REFERENCE 2 (bases 1 to 14568)

AUTHORS Storer, C.K., Pham, X.-O.T., Erwin, A.L., Mizoguchi, S.D., Warrenner, P., Hickey, M.J., Brinkman, F.S.L., Huftagle, W.O., Kowalik, D.J., Lagrou, M., Gardner, R.L., Goltz, L., Tolentino, E., Westbrock, Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Lardig, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Sailer, M.H., Hancock, R.E.W., Lory, S. and Olson, M.V.

TITLE Direct Submission

JOURNAL Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA

FEATURES

source

1. 14568  
Location/Qualifiers

organism="Pseudomonas aeruginosa"

strain="PA01"

/db\_xref="taxon:287"

83. .1033  
/gene="PA0148"

83. .1033  
/gene="PA0148"

83. .1033  
/gene="PA0148"

/codon\_start=1

/transl\_table=1

/product="probable adenosine deaminase"

/protein\_id="AG03538.1"

/db\_xref="GI:9945979"

/translation="MYEMLNALPKAEHLHLGTEPELLEFALAERNRIALFMDVET  
LRKAVYNNLOEFLDYAGADVRTEODPYDLTWYVLOKCAONVHVEPFEDQTH  
TDGRTPEVYALIRALRNGEKLIGIRGLIISFRLHSEDOAKLDOALPFRFA  
IANGDSSEYGHPPSKFQVFPDARSEGLTVAHAGEEPPEYIWEALDLLVERDH  
GVAFEDERLMRLIDEQIPLYVCPISNKLKCYFDDMSQTHIDMLERKVKTVNSD  
PAYFGVYTFENFHALQOSLGMEQEARLNLNSDLARLVK"

1231. .1776  
/gene="PA0149"

1231. .1776  
/gene="PA0149"

1231. .1776  
/gene="PA0149"

/codon\_start=1

/transl\_table=1

/product="probable sigma-70 factor, ECF subfamily"

/protein\_id="AG03539.1"

/db\_xref="GI:9945980"

/translation="MPADGEARRADLLKAFLOROMESLVSRYGCRATADLVQ  
LRLFRMRPAAPVEIGSYLLRSARNLADHLEGSROOLDMLPEOROGEPVGE  
QALBAGDQMRVREARLGRDPERTRRIFLLRHIGRTYAEIAQAMQLSQAIVEKHMRA  
LDACKASLAGPAEQARSAP"

1773. .2768  
/gene="PA0150"

1773. .2768  
/gene="PA0150"

1773. .2768  
/gene="PA0150"

/codon\_start=1

/transl\_table=1

/product="probable transmembrane sensor"

/protein\_id="AG03540.1"

/db\_xref="GI:9945981"

/translation="WTRPSDAAPYEPVLEQTAQAWLVRLGTPDATTARRAFEDWLAQ  
RPAPAPYARAEALMRLTEIPASRLAIEADALEYAPRRASRNPRRAYPLASVAC  
LLAMLMGMMQPCDMLQDMRDVYVAPGOVAEMKLSLIDAGSAVRLDGGO  
RRVLELGAALRYKRDALPFPVAGGRRVLTGEFVYRGAGATPVYTLQGRVVS  
AGTGAGAAEEVLAGQOVGFADKLGPPVADDAALAMRQGMISFTRRLAEYLDL  
ARYVPRILLDDALGRQPVSGFSRSDPEPAALKSLQAVLGRQILLGRLLVYIR"

2917. .5304  
/gene="PA0151"

2917. .5304  
/gene="PA0151"

2917. .5304  
/gene="PA0151"

/codon\_start=1

/transl\_table=1

/product="probable 10mb-dependent receptor"

/protein\_id="AG03541.1"

/db\_xref="GI:9945982"

/translation="WTPPPRRALTIPLALFISGLCIATVQAAEAVVRFSSQAAPLA  
QALNFSRASQTLITVETLPPQATAPALHSLVAEALQRLDLOGSLRRRIDERTL

TLPEIPASGALDATTVAAPQCYRPAANASISRGDPVLEIPQALISVPAQALQ  
DORPNLDDALNIGSTQANTIGTQDVAVMKRGSDNDSIMRGMSVQGRNETA  
TADHVEYKPAISLGIQDPGGVNVVWVSKPQLOQANALTLRGSAYAHNRGSGOL  
DSTGPDUSLSIALMLDHEDEPWRRDCKSRETIYAPSLAWGEEDTVNLSEHRE  
TTPREDRTADIPRTNRLDIPSRRLDEPNTITEGSELTITGLSETDLOAGMHDQ  
GYSRETYDDQARVAMVADGTLTRMDGTGHAVSSDSTTGLSDLOAGMHDQ  
FGMDHEKRIERADLLDSKTSREYLDVYGGVSTPVRVADSDQTKLSDSLTLE  
QDSHLDERLILVGARYQIYDQLAGRGREFNANDINQKWPVRGLVYRLSDEVSL  
YGSYRSFKNSNTIAPISGETIDSAIPEEATSWELGAKLDVPGRLSGTLAFIRK  
KNVYNELDCAAGNSSVAPAGRSRGIELDTGLQSEKSLISYSYAMLAEVDEPTL  
EGKRLQVAKRTSLASVYTLGCLFGDGRRLRGAGARVYGGKRGDPEANSEFDP  
DAFASDYTRREGGVKFLNWKMLFRTTYPSSANLKYAMGPRFOVSTYEE"

5681. .6508  
/gene="pca0"

5681. .6508  
/gene="pca0"

5681. .6508  
/gene="pca0"

/codon\_start=1

/transl\_table=1

/product="transcriptional regulator pca0"

/protein\_id="AG03542.1"

/db\_xref="GI:9945983"

/translation="MAVSOPAIKSTLKELELIGARLFERSKAGAEITENGVELNYA  
GPGVALRDGVNTIRGHPAOVEYRVGVLSVSGVYEVIRRLRHRRGALVSVANG  
PSAYLQALRGVGLDLYGRTDSDPDIQGLTEPELISESMALVYRGHPILAPLGLD  
QRLAEFLVPLPLACTTIRHADSLEFVGMGLAGQRLTSLPTLSRYVRSALIVYA  
PDAVRLDVNGCELAELDLGIREPGSVGLCHNAALPLPLGAQWFSVLEVAELH"

6643. .7362  
/gene="pcaH"

6643. .7362  
/gene="pcaH"

6643. .7362  
/gene="pcaH"

/codon\_start=1

/transl\_table=1

/product="3,4-dioxygenase, beta subunit"

/protein\_id="AG03543.1"

/db\_xref="GI:9945984"

/translation="MHDAESRRVYTRDRNMHPKALTPDYKTSVRSPOALVSPSL  
SETSGPDSHLRLGHRNDLLNFDHGLPGGKRIINRGVFDQYGRPVHTLVEMQA  
ANAGRYRHKNDRYLADLPNMGVGLDDESEHYFRITKPGYVWRGPNDMRPA  
HIHVSQSPALARLYVTQMFEGDPLIPKCPILITLADPEAAOSLIGRLDMSANPMD  
CLAYFVDVLRGORTFPENA"

7373. .7978  
/gene="pcaG"

7373. .7978  
/gene="pcaG"

7373. .7978  
/gene="pcaG"

/codon\_start=1

/transl\_table=1

/product="3,4-dioxygenase, alpha subunit"

/protein\_id="AG03544.1"

/db\_xref="GI:9945985"

/translation="MSHELRLREPSQTAGPYVHIGLALAAAGNTPRDLLETNRARRG  
ARGOHTLLGHVYDNGHLYVSDSLEFQWQDSSEGYQORYPERPNSFGRTATPDA  
GFWTLHTVYKPGVNRAGVPAPAHINVSLEFARGINIHQRTLYEDDETEANACPVIN  
ILEQAPRDRITLAQCEVDQALAYREFDIRIOGNETVFDE"

8185. .9024  
/gene="pcaR"

8185. .9024  
/gene="pcaR"

8185. .9024  
/gene="pcaR"

/codon\_start=1

/transl\_table=1

/product="transcriptional regulator pcar"

/protein\_id="AG03545.1"

/db\_xref="GI:9945986"

/translation="MSELPHEPATLAPPTVSPARKIAFTGDPNMTSLARGLAVIH  
AFOERKRLTIAQISHREIIPRAAVRCHITMOGVTATGCRYSLLPKVTLGHAY  
ISSTPLATAPQILDRISRLHACSMATLEEDVYIARSATPQRLISVDNLGSHY  
PAYCTSMGILLALDDDAITAHVPGGVQAKTSRTLYTPETILCLIVEIRROGCIV  
DELEVEGLRSIAVPVDSAGHVALALNVTGTHAGRVSRADIESRFLPLLEASRELSAR  
LPE"

9177. .10328  
/gene="pcaR"

```

CDS
/gene="PA0156"
/177..10328
/gene="PA0156"
/codon_start=1
/transl_table=11
/product="probable RND efflux membrane fusion protein precursor"
/protein_id="AAG03546.1"
/db_xref="GI:9945987"
/translation="MSDARGAFHSKGRMSRMLPAALICAGLHVGCGAAPPAAEHVRL
AQTVMKAEFASATITGDIQARVADOSFRVGGKIVERLVGVGHVAAQVLAFLDPO
DORSNENQAQAAVAAQAAQSLADLNTYORAKLLPKVTSQSEYDQALASVRSQSS
KAAQALANARDLSTYELRASDAGVITARQAEVGOVQAVPFTTLARQGBRDVFN
VYSELSHDVDCORITVSLKPEVTASCKVREITPPYDESGTILKRVGDSVPAEM
SLGSVYMASVAPAEHSVITPMSATSKGEOPAWLLDQOKARLQVRAVARSERY
VIDGLEGAGOTVYVYGGDLHFGVVEVADPPQTSASRDVAGGQP"
10325..11395
/gene="PA0157"
/10325..11395
/gene="PA0157"
/codon_start=1
/transl_table=11
/product="probable RND efflux membrane fusion protein precursor"
/protein_id="AAG03547.1"
/db_xref="GI:9945988"
/translation="MKPFSIAGLGFALLSGCDEDEPPAPPPVLTMYTKTKNDL
GRFASIQARVESVYGFRTNGRIASRLFDGDEGKALLATLDPDQOQILASQSD
LASAEQILDAQANAROEELFARSVAOARLDARPLRTSASDQAKAAVQQRD
QLSYRLVTFDFDGIITVWMAEGOVVAGAVVTLARPEKREAFDLPVAAVSLPAD
ARPLVSAQLDPOARTGTSIRLGRQADASTRIKRVRLSLAQTEARLSTIOVQLSS
AGRSVTPASVQLERDGTQVWVDKSSVALREVQVLSRDERQVVGQGLADDR
VVRAGVNSLAKGQIKLDEDA"
11392..14439
/gene="PA0158"
/11392..14439
/gene="PA0158"
/11392..14439

Query Match
Best Local Similarity 55.6% Pred. No. 0.4;
Matches 80; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 63 GTCACCTTTCACAGCCAGGTGATGACATGAAGAATCTGTATGATGATCACAACAC 122
DB 12484 GTGGCCTCTGCATCCCGCTGCTGCGCATGTCTTCTGCTTCATGATGATCACC 12543
QY 123 AACACCATTCAAATGTTACCGCAGCAGCAGCCCAACAACACAGGCTGTGAAGCC 182
DB 12544 ATCACCATGCGAGCGGGTTTCCCTCGCGCGCTGATCATCGCCCTCGGCGCTGTG 12603
QY 183 AATTTCCTGAACGTCCAGGATATG 206
DB 12604 GATGCATGATCACCCTGGAGATG 12627

RESULT 14
AF128411S6 1077 bp DNA linear PRI 02-JAN-2000
DEFINITION Homo sapiens carbonic anhydrase VI gene, exon 6.
ACCESSION AF128416
VERSION AF128416.1 GI:6652972
KEYWORDS
SEGMENT
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1077)
REFERENCE
AUTHORS Grubb, D.J.
TITLE Direct Submision
JOURNAL Submitted (16-FEB-1999) Howard Florey Institute, University of
Melbourne, Royal Parade, Melbourne, Vic 3133, Australia
FEATURES
source
1..1077
Location/Qualifiers

```

```

BASE COUNT 239 a 309 c 228 g 301 t
ORIGIN
exon
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1p36.2"
/number=6

Query Match
Best Local Similarity 72.0% Pred. No. 0.33;
Matches 54; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 15 ACACCCCCCGCAGACAGATGCCAGTGGTGGCTTACGACAAAGTCACTTGTCC 74
DB 606 ACCGCTCCCGCAGCTAGACAGTCCAGTGGTGGCTGGCAGATTTGCAAGCTCC 665
QY 75 AAGCCCGAGGTGGT 89
DB 666 AGGACACAGGTATG 680

RESULT 15
AC098147 82576 bp DNA linear HTG 21-DEC-2001
LOCUS AC098147
DEFINITION Rattus norvegicus clone CH230-173010, *** SEQUENCING IN PROGRESS -
ACCESSION AC098147
VERSION AC098147.2 GI:17975713
KEYWORDS HTG; HTGS; PHASEL.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 82576)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaralunga, H.C., Aze, J.R., Banks, T., Barbara, J.,
Beaton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J.,
Bowie, S., Brieve, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C.,
Butcher, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hognes, M., Holloway, C.,
Hollins, B., Homsl, F., Howard, S., Huber, J., Huiyk, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewiss, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Lounseged, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M.,
Mel, G., Metzker, M., Miner, G., Moser, M., Neal, D., Newton, J., Newton, N.,
Morgan, A., Morris, S., Moser, M., Moser, M., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwenkwo, S.,
Ogih, M., Okunolu, G., Oragunye, N., Oyedro, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.,
Ruit, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N.,
Sisson, I., Sodergren, E., Sonakke, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Teiford, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Wasyngton, C.,
Watlington, S., Williams, G., Williamson, A., Wlezyk, R., Wooden, S.,

```



## FEATURES

**Source**

Location/Qualifiers  
1 82576

1. 82576  
/Corran.com

```

/organism="Rattus norvegicus"
/db_xref="taxon:10116"

```

```
/db_xref="taxon:10116"
```

```
/clone="CH230-173010"
```

Query match  
Best Local

```
7.68; Score 40.6; DB 2; Length 82576;
19.38; Prod No 0 00.
```

Best Local Similarity 49.3%; Pred. NO. 0.99;  
Matches 106; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

|         |      |              |    |            |      |        |    |      |    |
|---------|------|--------------|----|------------|------|--------|----|------|----|
| Matches | 106; | Conservative | 0; | Mismatches | 109; | Indels | 0; | Gaps | 0; |
|---------|------|--------------|----|------------|------|--------|----|------|----|

95 AGAAACTCTGTATGGATCACAACAACACCATTCAAATGGTTACCGCAGCACACA 154

Db 49949 AAACACATTTTAGGATATCATGAACAATTTTATCTCAAAAATGAGAAGAACCATTTGT 50008

155 GCCCAACACACAGGGTGTGGAAGCCAATTCTCTGAACGTCACGATATGTA  
214

Db 50009 TTTTCAACAATGGGAGATAAAAAAGCAGCTGCAAGGTTGGGATTAGCTCAGTG 50068

215 GTACCACCTTTACCTAAAAACATGCAGAAGGAGATTCTACAACCTAAGAAACAGAAGAA 274

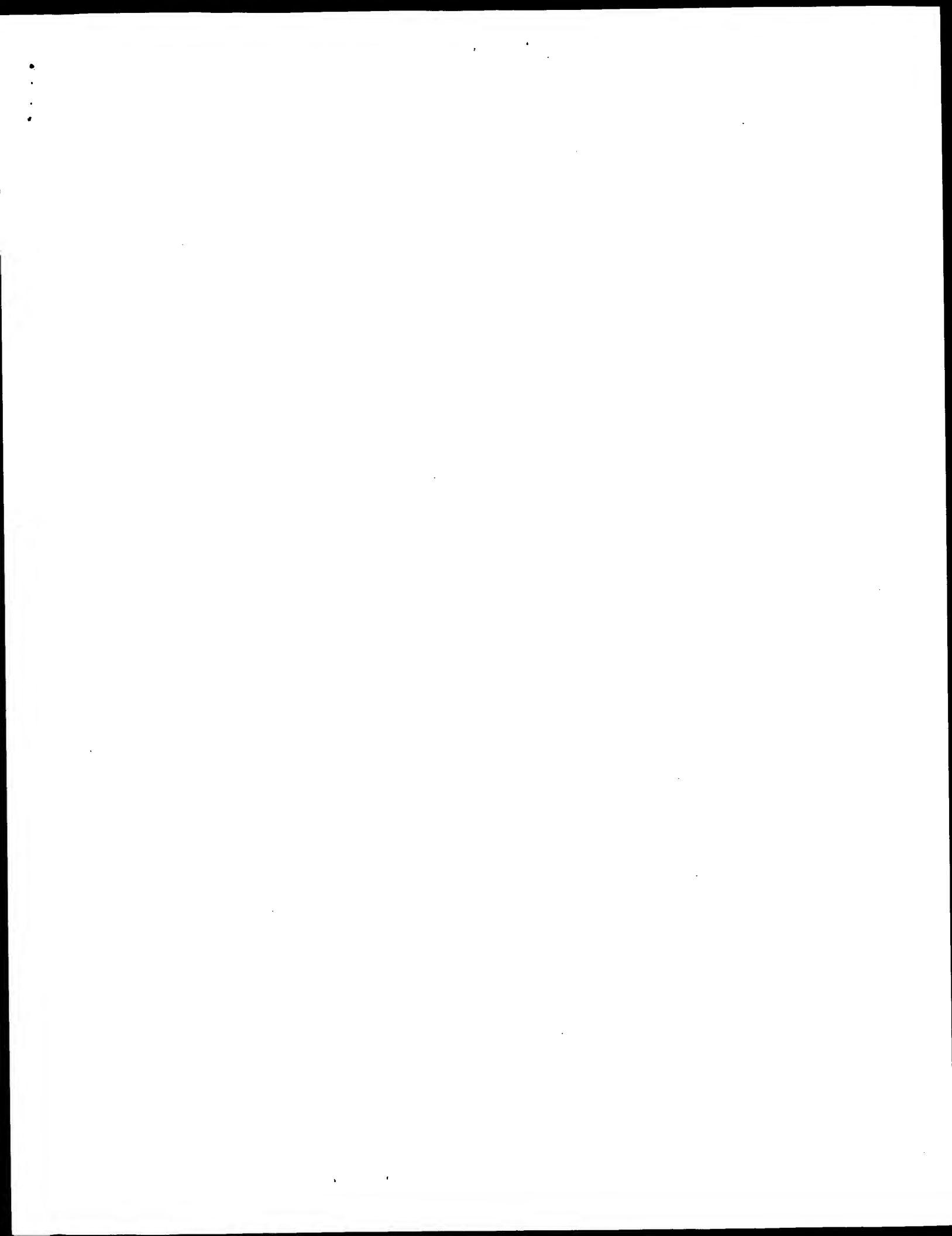
Db 50069 GTAGAACGCTTGCCCTACGACGACAAGGCCCTGGATTCCGTCCTCCAGCTCAAAAAAAAAA 50128

275 AACAAAGAGAACC GGCACTTTTGAGACCGGAAT 309

Db 50129 AAAAAAAAAAGCAGCTGCACCTTTT TAGACCTCAAT 50163

Search completed: October 19, 2002, 06:35:17  
Job time : 2448 secs

Job time : 2448 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 19, 2002, 04:30:01 ; Search time 201 Seconds

(without alignments)  
4544.271 Million cell updates/sec

Title: AI327498

Perfect score: 532

Sequence: 1 AATTCGGCAGTACAGACACC.....CATTAATTCCTGCATCAAT 532

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

|     |   |
|-----|---|
| 1:  | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT.*  |
| 2:  | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT.*  |
| 3:  | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT.*  |
| 4:  | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT.*  |
| 5:  | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT.*  |
| 6:  | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT.*  |
| 7:  | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT.*  |
| 8:  | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT.*  |
| 9:  | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT.*  |
| 10: | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT.*  |
| 11: | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT.*  |
| 12: | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT.*  |
| 13: | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT.*  |
| 14: | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT.*  |
| 15: | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT.*  |
| 16: | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT.*  |
| 17: | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT.*  |
| 18: | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT.*  |
| 19: | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT.*  |
| 20: | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT.*  |
| 21: | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT.*  |
| 22: | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.* |
| 23: | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT.* |
| 24: | /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT.*  |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1          | 38.8  | 7.3         | 5688   | 23    | ABL26543    |
| 2          | 38.8  | 7.3         | 7588   | 23    | ABL26542    |
| 3          | 38    | 7.1         | 13631  | 23    | ABL20354    |
| 4          | 37    | 7.0         | 400    | 23    | AA575460    |
| 5          | 37    | 7.0         | 963    | 23    | AA568580    |
| 6          | 35.8  | 6.7         | 12507  | 24    | ABL32298    |
| 7          | 35.6  | 6.7         | 51     | 22    | ABL33398    |
| 8          | 35.2  | 6.6         | 364    | 22    | AA184802    |
| 9          | 35.2  | 6.6         | 966    | 21    | AA42768     |

|    |      |     |       |    |          |                      |
|----|------|-----|-------|----|----------|----------------------|
| 10 | 35.2 | 6.6 | 1235  | 21 | AA50838  | Arabidopsis thaliana |
| 11 | 35.2 | 6.6 | 1239  | 21 | AA53292  | Arabidopsis thaliana |
| 12 | 34.2 | 6.4 | 496   | 22 | AA191273 | Human polynucleoti   |
| 13 | 34.2 | 6.4 | 2022  | 22 | AAH84529 | E. coli growth and   |
| 14 | 33.8 | 6.4 | 663   | 23 | AA565332 | DNA encoding novel   |
| 15 | 33.8 | 6.4 | 663   | 23 | AA580842 | DNA encoding novel   |
| 16 | 33.8 | 6.4 | 793   | 22 | AA195022 | Human neuroblastom   |
| 17 | 33.8 | 6.4 | 5163  | 21 | AAV20700 | Cryptosporidium pa   |
| 18 | 33.8 | 6.4 | 5163  | 21 | AAV61849 | ORF encoding a por   |
| 19 | 33.8 | 6.4 | 5318  | 21 | AAA61848 | DNA encoding a por   |
| 20 | 33.6 | 6.3 | 1518  | 23 | AA553583 | Helicobacter pylor   |
| 21 | 33.6 | 6.3 | 5318  | 19 | AAV20701 | Cryptosporidium pa   |
| 22 | 33.6 | 6.3 | 7286  | 20 | AA57224  | WO 9923223 Seq ID    |
| 23 | 33.6 | 6.3 | 7938  | 20 | AA57235  | WO 9923223 Seq ID    |
| 24 | 33.4 | 6.3 | 466   | 22 | AA182085 | Human immune/haema   |
| 25 | 33.4 | 6.3 | 617   | 22 | AAK71856 | Human immune/haema   |
| 26 | 33.4 | 6.3 | 941   | 22 | AAK71857 | Human immune/haema   |
| 27 | 33.4 | 6.3 | 9516  | 22 | AA542103 | Genomic sequence #   |
| 28 | 33.2 | 6.2 | 5181  | 23 | ABL10975 | Drosophila melanog   |
| 29 | 33.2 | 6.2 | 14974 | 23 | ABL10974 | Drosophila melanog   |
| 30 | 33   | 6.2 | 441   | 23 | AA575453 | DNA encoding novel   |
| 31 | 33   | 6.2 | 948   | 23 | ABL03799 | Drosophila melanog   |
| 32 | 33   | 6.2 | 2067  | 21 | AA550488 | soybean sulphate p   |
| 33 | 33   | 6.2 | 3399  | 17 | AA705868 | Chicken leucocytoz   |
| 34 | 33   | 6.2 | 5813  | 24 | ABL33804 | Human immune syste   |
| 35 | 33   | 6.2 | 6151  | 24 | ABL33611 | Human immune syste   |
| 36 | 33   | 6.2 | 8305  | 24 | ABL33569 | Human immune syste   |
| 37 | 32.8 | 6.2 | 1586  | 16 | AA087587 | DNA encoding leuco   |
| 38 | 32.8 | 6.2 | 4590  | 22 | AAH24065 | Chemically treated   |
| 39 | 32.8 | 6.2 | 6071  | 24 | AAAD2316 | Human immune syste   |
| 40 | 32.8 | 6.2 | 6853  | 24 | ABL32311 | Human immune syste   |
| 41 | 32.8 | 6.2 | 8446  | 24 | ABL32311 | Human immune syste   |
| 42 | 32.8 | 6.2 | 10877 | 21 | AAZ55620 | Human BMP-7 gene 5   |
| 43 | 32.6 | 6.1 | 5014  | 22 | AAV58286 | S. pyogenes SP-14-   |
| 44 | 32.4 | 6.1 | 357   | 22 | AA187311 | Human polynucleoti   |
| 45 | 32.4 | 6.1 | 1664  | 20 | AA28072  | Heparin sulphate 6   |

#### ALIGNMENTS

RESULT 1  
ID ABL26543  
ABL26543 standard; DNA; 5688 Bp.

XX ABL26543;  
AC 26-MAR-2002 (first entry)  
XX  
XX  
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 31102.  
DE  
XX  
XX Drosophila: developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX  
XX Drosophila melanogaster.  
OS  
XX  
XX WO200171042-A2.  
PN  
XX  
XX 27-SEP-2001.  
PD  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
PF  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR  
XX  
XX 11-JUL-2000; 2000US-0614150.  
PA  
XX  
XX (PEKE ) PE CORP NY.  
PI  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW,  
DR  
XX  
XX WPI; 2001-656860/75.  
PT  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
genes from Drosophila and for elucidating cell signalling and cell-cell



|                          |   |
|--------------------------|---|
| PT                       | interactions -  |
| XX                       |   |
| PS                       | Claim 1; SEQ ID NO 31102; 21pp + Sequence Listing; English.               |
| CC                       |   |
| CC                       | The invention relates to an isolated nucleic acid detection reagent       |
| CC                       | capable of detecting 1000 or more genes from Drosophila. The invention is |
| CC                       | useful in developmental biology and in elucidating cell signalling and    |
| CC                       | cell-cell interactions in higher eukaryotes for the development of        |
| CC                       | insecticides, therapeutics and pharmaceutical drugs. The invention        |
| CC                       | discloses genomic DNA sequences (AB16176-AB130511), expressed DNA         |
| CC                       | sequences (AB101840-AB16175) and the encoded proteins                     |
| CC                       | (ABB57737-ABB72072).  |
| CC                       | The sequence data for this patent did not form part of the printed        |
| CC                       | specification, but was obtained in electronic format directly from WIPO   |
| CC                       | at ftp.wipo.int/pub/published_pct_sequences.                              |
| XX                       |   |
| SEQ                      | Sequence 5688 BP; 1925 A; 1719 C; 1147 G; 897 T; 0 other;                 |
|                          |   |
| Query Match              | 7.3%; Score 38.8; DB 23; Length 5688;                                     |
| Best Local Similarity    | 54.1%; Pred. No. 0.35;  |
| Matches 79; Conservative | 0; Mismatches 67; Indels 0; Gaps 0  |
|                          |   |
| QY                       | 115 ACACACAAACACCCATTCAAATGGTTACCGCAGCAGCAGCCCAACCAACAGGGTGG 174          |
| DB                       | 5519 ACACACGAGAACATCCAAAGCAAGATCAGAACCAAGAACACATGCAAAAGCAGA 5578          |
| QY                       | 175 TGGAAAGCCATTTCTCGAAGCTCCAGATATGTACTCTTCGTACCACTTTACCTAAAA 234         |
| DB                       | 5579 ATCAGAACAAAAACCGAAGCATCCAAAGCAAGATCAGAACCAACAGAACAAAGACA 5638        |
| QY                       | 235 ACATGCGAAGGAGATTCTACCAACT 260   |
| DB                       | 5639 ATGAGCGAAGAAAGAAACCCAACT 5664  |
|                          |   |
| RESULT 2                 |   |
| ABL26542/c               |   |
| ID                       | ABL26542 standard; DNA; 7588 BP.  |
| XX                       |   |
| AC                       | ABL26542;   |
| XX                       |   |
| DT                       | 26-MAR-2002 (first entry)   |
| XX                       |   |
| DE                       | Drosophila melanogaster genomic polynucleotide SEQ ID NO 31099.           |
| XX                       |   |
| KW                       | Drosophila; developmental biology; cell signalling; insecticide;          |
| XX                       | pharmaceutical; gene; ds.   |
| XX                       |   |
| OS                       | Drosophila melanogaster.  |
| XX                       |   |
| PN                       | WO200171042-A2.   |
| XX                       |   |
| PD                       | 27-SEP-2001.  |
| XX                       |   |
| PF                       | 23-MAR-2001; 2001WO-US09231.  |
| XX                       |   |
| PR                       | 11-MAR-2000; 2000US-191637P.  |
| XX                       |   |
| PR                       | 11-JUL-2000; 2000US-0614150.  |
| XX                       |   |
| PA                       | (PEKE ) PE CORP NY.   |
| XX                       |   |
| PI                       | Venter JC, Adams M, Li PWD, Myers EW;                                     |
| XX                       |   |
| DR                       | WPI; 2001-656860/75.  |
| XX                       |   |
| PT                       | New isolated nucleic acid detection reagent for detecting 1000 or more    |
| PT                       | genes from Drosophila and for elucidating cell signalling and cell-cell   |
| PT                       | interactions -  |
| XX                       |   |
| PS                       | Claim 1; SEQ ID NO 31099; 21pp + Sequence Listing; English.               |
| XX                       |   |
| CC                       | The invention relates to an isolated nucleic acid detection reagent       |
| CC                       | capable of detecting 1000 or more genes from Drosophila. The invention is |

|   |   |
|---|---|
| CC  | useful in developmental biology and in elucidating cell signalling and  |
| CC  | cell-cell interactions in higher eukaryotes for the development of      |
| CC  | insecticides, therapeutics and pharmaceutical drugs. The invention      |
| CC  | discloses genomic DNA sequences (AB16176-AB130511), expressed DNA       |
| CC  | sequences (AB101840-AB16175) and the encoded proteins                   |
| CC  | (ABB57737-ABB72072).  |
| CC  | The sequence data for this patent did not form part of the printed      |
| CC  | specification, but was obtained in electronic format directly from WIPO |
| CC  | at ftp.wipo.int/pub/published_pct_sequences.                            |
| SQ  | Sequence 7588 BP; 1405 A; 1605 C; 2192 G; 2386 T; 0 other;              |
| Query Match   | 7.3%; Score 38.8; DB 23; Length 7588;                                   |
| Best Local Similarity   | 54.1%; Pred.No. 0.39; Mismatches 67; Indels 0; Gaps 0;                  |
| Matches 79; Conservative  | 0; Mismatches 67; Indels 0; Gaps 0;                                     |
| Dy  | 115 ACACAGCAACACCATTCAAATGTTTACCGCAGCACACGCCAACAACACAGGGGTG 174         |
| Dd  | 886 ACAACCCGAAGAACATCTCCAAAGCAAGAAATCAGAACACCATGCGAAAGCA 827            |
| Dy  | 175 TGGAAAGCAATTTCTCGAAGTCGAGATATGTACTCTTGCTACCACTTTACTATAAA 234        |
| Dd  | 826 ATCAGAACAACAAAACCGAAGATCTCAAGCAAGCAATCAGAACACAGAACAAAGACA 767       |
| Oy  | 235 ACATGCAGAAAGAGATTTCTCAACCT 260                                      |
| Dd  | 766 ATGAGCAGAAAGAAACCAACCT 741  |
| RESULT 3  |   |
| ABL20354  |   |
| ID  | ABL20354 standard; DNA: 13631 BP.                                       |
| XX  | ABL20354;   |
| AC  |   |
| DF  |   |
| DE  | 26-MAR-2002 (first entry)   |
| XX  |   |
| XX  | Drosophila melanogaster genomic polynucleotide SEQ ID NO 12535.         |
| KW  | Drosophila; developmental biology; cell signalling; insecticide;        |
| OS  | pharmaceutical; gene; ds.   |
| PN  | Drosophila melanogaster.  |
| WO200171042-A2.   |   |
| 27-SEP-2001.  |   |
| 23-MAR-2001; 2001WO-US09231.  |   |
| 23-MAR-2000; 2000US-191637P.  |   |
| 11-JUL-2000; 2000US-061415O.  |   |
| (PEKE ) PE CORP NY.   |   |
| Venter JC, Adams M, Li PWD, Myers EW;                                     |   |
| WPI, 2001-656860/75.  |   |
| New isolated nucleic acid detection reagent for detecting 1000 or more    |   |
| genes from Drosophila and for elucidating cell signalling and cell-cell   |   |
| interactions -  |   |
| Claim 1; SEQ ID NO 12535; 21bp + Sequence Listing; English.               |   |
| The invention relates to an isolated nucleic acid detection reagent       |   |
| capable of detecting 1000 or more genes from Drosophila. The invention is |   |
| useful in developmental biology and in elucidating cell signalling and    |   |
| cell-cell interactions in higher eukaryotes for the development of        |   |
| insecticides, therapeutics and pharmaceutical drugs. The invention        |   |
| discloses genomic DNA sequences (AB16176-AB130511), expressed DNA         |   |
| sequences (AB101840-AB16175) and the encoded proteins                     |   |
| (ABB57737-ABB72072).  |   |



Sequence 963 BP; 353 A; 171 C; 268 G; 169 T; 2 other

|     |  |     |
|-----|--|-----|
| 231 | AAAAACATGCAGAGCAGATTTCTACAACCTAAGAAACGAAAGAAAAACAAGAAAGACCGG | 239 |
| QY  |  |     |
| 257 | TTCTGACGAGAGAGAGAGAGAGAGAAANNAGAGAGAGAGAGAGAGAGAGAGAGAGAG    | 316 |

[illegible]

RESULT 6  
ABL32298/c  
ID ABL32298 standard; DNA; 12507 BP

-XX  
AC ABL32298

|    |             |               |
|----|-------------|---------------|
| XX | 26-MAR-2002 | (first entry) |
| DT |             |               |

| XX | Human immune system associated gene | SEQ ID NO: | 271 |
|----|-------------------------------------|------------|-----|
| DE |                                     |            |     |
| XX |                                     |            |     |

KW Human, immune system disease; cytosine methylation; antiscclerotic  
 KW antiarteriosclerotic; antihaemic; cytosolic; noctropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antinflammatory; cancer; eye disease; arteriosclerosis; anemia;  
 KW acute myeloid leukemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease  
 KW gene; ds.

|    |                |
|----|----------------|
| XX | Homo sapiens.  |
| OS |                |
| XX |                |
| PN | W0200200928-A2 |

XX  
PD 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.  
PF  
JUL

|    |                             |
|----|-----------------------------|
| XX | 30-JUN-2000; 2000DE-1032529 |
| PR | 01-SEP-2000; 2000DE-1043826 |
| PP |                             |

PA (EPIC-) EPIGENOMICS AG

XX Olek A, Piepenbrock C, Berlin K;  
PI

XX WPI; 2002-130909/17.  
DR

XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT and treatment of diseases associated with abnormal

PT Iordianosis and leucemia or carcinoma  
PT cytosine methylation -

XX  
PS  
XX

Query Match 6.7%; Score 35.8; DB 24; Length 12507

| Best Local Matches | Similarity | Pred. No. 4; | Indels | Gaps |
|--------------------|------------|--------------|--------|------|
| 97; Conservative   | 48.7%      | 0;           | 102;   | 0;   |
| Mismatches 147     |            |              |        |      |

88 TGACCATAGAAAACCTCTGTTATGATCACAACAACAACACCATCAATCAAAATGGTTCACCA 35

Db 3613 TAAATACACCTAAATATAATTAACATAATCCCAATCCCAAGATATGT 20

[illegible]

D6 3553 CAAATTAAGTAAAGGACCTAATCGAAGCAGGATTTCAACCTAGAAGAAC 267

308 TACTGTGGTACCACCTTTACCTAAAAACATGCAGAGAGATTTCAACCTAGAAGAAC 267

DB 3493 ACACAAACACACATTAATTCACAAACAAAAATTAATAAAATTTCCCAACCAATTAATA 3434

268 AGAAGAAACCAAGAAGAA 286

Db 3433 AAAAAAAAAAAAAAAAAA 3415

RESULT 7  
AAL33398

XX ID AAL33398 SLAIDAO, DNO, 2-1-1988

AC AAL55530,  
XX  
DT 2A-JAN-2002 (first entry)

|    |                                 |
|----|---------------------------------|
| DI |                                 |
| XX | Human SNP oligonucleotide #6606 |
| DE |                                 |

XX Immunosuppressive; immunostimulatory; antinflammatory; cytostatic;  
KM antineoplastic; gene therapy; vaccine; amyase; cancer;

KW neuroprotective; antimicrobial; gene therapy  
 amyloid protein; angiotensin; apoptosis related protein; cadherin;  
 histone; kinase; colony stimulating factor

KM cyclin; polymerase; oncogene; interferon; complement related protein; cytochrome; kinesin; cytokine; interleukin; 14-3-3; G-protein coupled receptor; thioesterase; inflammation

KW multifactorial disease; autoimmune disease; infection;  
interleukin, 6 protein; protein, 6  
KW nervous system disease; ss.  
KW multifactorial disease; ss.

|    |               |
|----|---------------|
| NM | Homo sapiens. |
| XX |               |
| OS |               |

WO200147944-A2

XX  
PD 05-JUL-2001

XX  
PF 28-DEC-2000; 2000WO-US35498.

|    |              |                |
|----|--------------|----------------|
| XX | 28-DEC-1999; | 99US-0173419   |
| PR | 27-DEC-2000. | 2000US-0173419 |

PR 2 / DEC 2000, 20000222  
XX  
PA (CUBA-) CUBAGEN CORP.

FA (Comm) 100-100000  
XX  
PI Shinkets RA, Leach N

XX  
DR WPI; 2001-465210/50.

XX polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases  
PT for diagnosis and treating, e.g.

PT oncogenes and histones, useful for diagnosis of cancer, autoimmune diseases and infections

|    |                                     |
|----|-------------------------------------|
| XX | Claim 1; Page 3269; 4143pp; English |
| PS |                                     |

XX The present invention relates to oligonucleotides encoding polymorphic  
CC variants of proteins related to anylases, amyloid proteins, angiotensin,  
CC apoptosis related proteins, cadherin, cyclin, polymerase, oncogenes,  
CC histones, kinases, colony stimulating factors, complement related  
CC proteins, cytochromes, kinesins, cytokines, interferons, interleukins,  
CC G-protein coupled receptors and thioesterases. The present sequence is  
CC one such oligonucleotide. The oligonucleotides and the peptides encoded  
CC by them may be used in the prevention, diagnosis and treatment of  
CC diseases associated with inappropriate expression of the proteins listed  
CC above. Disorders that may be prevented, diagnosed and/or treated include  
CC multifactorial diseases with a genetic component, such as autoimmune  
CC diseases (e.g. rheumatoid arthritis, multiple sclerosis, diabetes,  
CC systemic lupus erythematosus and Grave's disease), inflammation, cancer  
CC (e.g. cancers of the bladder, brain, breast, colon and kidney,  
CC leukemia), diseases of the nervous system and an infection of pathogenic  
CC organisms.

XX Sequence 51 BP; 15 A; 17 C; 11 G; 8 T; 0 other;

XX Query Match 6.7%; Score 35.6; DB 22; Length 51;  
Best Local Similarity 82.0%; Pred. No. 0.39;  
Matches 41; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 140 TTACCGACGACACACCCCAACACACAGGCTGTGGTGAAGCAATTTC 189  
DB 2 TTACCGACGACACACCCCAACACACAGGCTGTGGTGAAGCAATTTC 51

RESULT 8  
AA184802  
ID AA184802 standard; cDNA; 364 BP.  
XX AA184802;  
AC  
XX 06-NOV-2001 (first entry)  
DT  
XX Human polynucleotide SEQ ID NO 4862.  
DE  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation; ss.  
OS Homo sapiens.  
XX  
XX MO200164835-A2.  
PN  
XX 07-SEP-2001.  
PD  
XX 26-FEB-2001; 2001WO-US04927.  
PE  
XX 28-FEB-2000; 2000US-0515126.  
PR  
XX 18-MAY-2000; 2000US-0577409.  
PX  
XX (HYSE-) HYSEQ INC.  
PA  
XX Tang YT, Liu C, Drmanac RT;  
PI  
XX MPI: 2001-514838/56.  
DR  
XX P-PSDB: AAO04871.  
PT  
XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX  
XX Claim 1; SEQ ID NO 4862; 1399bp + Sequence Listing; English.  
PS  
XX The invention relates to human polynucleotides (AA179941-AA193841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or

CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pat\_sequences.

XX Sequence 364 BP; 172 A; 58 C; 71 G; 63 T; 0 other;

XX Query Match 6.6%; Score 35.2; DB 22; Length 364;  
Best Local Similarity 53.7%; Pred. No. 1.2;  
Matches 73; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 151 CACAGCCCAACACACAGGCTGTGGTGAAGCAATTTCCTGAAGCTCCAGATATGACT 210  
DB 1 CCCAGCGCTCCGCCACAGGCTGTCCGATCGGCCCATATCCCGAAATGTGTATACCT 60

QY 211 CTTCGTACCACTTTACTTAAAAACATGACAGAGAGATCTCAACCTAAGAACAGA 270  
DB 61 TCCCGTACTTACCAAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 120

QY 271 AGAAAAACAAGAAGA 286  
DB 121 AAAAAAAGAAAAAAGAA 136

RESULT 9  
AAC42768  
ID AAC42768 standard; DNA; 966 BP.  
XX AAC42768;  
AC  
XX 17-OCT-2000 (first entry)  
DT  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 36776.  
DE  
XX  
XX Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
XX Arabidopsis thaliana.  
OS  
XX  
XX EP1033405-A2.  
PN  
XX 06-SEP-2000.  
PD  
XX 25-FEB-2000; 2000EP-0301439.  
PE  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134270.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 22-JUN-1999; 99US-0139817.  
PR 23-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 28-JUN-1999; 99US-0140695.  
PR 29-JUN-1999; 99US-0140823.  
PR 30-JUN-1999; 99US-0140991.  
PR 01-JUL-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149175.  
PR 17-AUG-1999; 99US-0149426.  
PR 18-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 01-SEP-1999; 99US-0151438.  
PR 07-SEP-1999; 99US-0151930.  
PR 10-SEP-1999; 99US-0152363.  
PR 13-SEP-1999; 99US-0153070.  
PR 15-SEP-1999; 99US-0153758.  
PR 16-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.

|    |              |              |
|----|--------------|--------------|
| PR | 07-MAY-1999; | 9905-0132863 |
| PR | 11-MAY-1999; | 9905-0134216 |
| PR | 14-MAY-1999; | 9905-0134218 |
| PR | 14-MAY-1999; | 9905-0134221 |
| PR | 14-MAY-1999; | 9905-0134221 |
| PR | 14-MAY-1999; | 9905-0134760 |
| PR | 18-MAY-1999; | 9905-0134378 |
| PR | 19-MAY-1999; | 9905-0134641 |
| PR | 20-MAY-1999; | 9905-0135124 |
| PR | 21-MAY-1999; | 9905-0135253 |
| PR | 24-MAY-1999; | 9905-0135529 |
| PR | 25-MAY-1999; | 9905-0136021 |
| PR | 27-MAY-1999; | 9905-0136392 |
| PR | 28-MAY-1999; | 9905-0136782 |
| PR | 01-JUN-1999; | 9905-0137222 |
| PR | 03-JUN-1999; | 9905-0137528 |
| PR | 04-JUN-1999; | 9905-0137502 |
| PR | 07-JUN-1999; | 9905-0137824 |
| PR | 08-JUN-1999; | 9905-0138094 |
| PR | 10-JUN-1999; | 9905-0138540 |
| PR | 10-JUN-1999; | 9905-0138847 |
| PR | 16-JUN-1999; | 9905-0139119 |
| PR | 16-JUN-1999; | 9905-0139452 |
| PR | 17-JUN-1999; | 9905-0139452 |
| PR | 18-JUN-1999; | 9905-0139454 |
| PR | 18-JUN-1999; | 9905-0139455 |
| PR | 18-JUN-1999; | 9905-0139456 |
| PR | 18-JUN-1999; | 9905-0139457 |
| PR | 18-JUN-1999; | 9905-0139458 |
| PR | 18-JUN-1999; | 9905-0139460 |
| PR | 18-JUN-1999; | 9905-0139461 |
| PR | 18-JUN-1999; | 9905-0139462 |
| PR | 18-JUN-1999; | 9905-0139463 |
| PR | 18-JUN-1999; | 9905-0139617 |
| PR | 22-JUN-1999; | 9905-0139819 |
| PR | 22-JUN-1999; | 9905-0139750 |
| PR | 23-JUN-1999; | 9905-0139750 |
| PR | 23-JUN-1999; | 9905-0140353 |
| PR | 24-JUN-1999; | 9905-0140695 |
| PR | 28-JUN-1999; | 9905-0140821 |
| PR | 29-JUN-1999; | 9905-0140991 |
| PR | 30-JUN-1999; | 9905-0141287 |
| PR | 01-JUL-1999; | 9905-0141842 |
| PR | 02-JUL-1999; | 9905-0142154 |
| PR | 06-JUL-1999; | 9905-0142055 |
| PR | 08-JUL-1999; | 9905-0142380 |
| PR | 09-JUL-1999; | 9905-0142920 |
| PR | 13-JUL-1999; | 9905-0142977 |
| PR | 15-JUL-1999; | 9905-0143542 |
| PR | 16-JUL-1999; | 9905-0143654 |
| PR | 16-JUL-1999; | 9905-0144005 |
| PR | 19-JUL-1999; | 9905-0144086 |
| PR | 19-JUL-1999; | 9905-0144325 |
| PR | 19-JUL-1999; | 9905-0144331 |
| PR | 19-JUL-1999; | 9905-0144332 |
| PR | 19-JUL-1999; | 9905-0144334 |
| PR | 20-JUL-1999; | 9905-0144335 |
| PR | 20-JUL-1999; | 9905-0144352 |
| PR | 21-JUL-1999; | 9905-0144384 |
| PR | 21-JUL-1999; | 9905-0144814 |
| PR | 22-JUL-1999; | 9905-0145086 |
| PR | 22-JUL-1999; | 9905-0145088 |
| PR | 22-JUL-1999; | 9905-0145087 |
| PR | 22-JUL-1999; | 9905-0145088 |
| PR | 22-JUL-1999; | 9905-0145089 |
| PR | 22-JUL-1999; | 9905-0145192 |

|    |             |               |
|----|-------------|---------------|
| PR | 23-JUL-1999 | 99US-01455145 |
| PR | 23-JUL-1999 | 99US-01455218 |
| PR | 23-JUL-1999 | 99US-01455224 |
| PR | 26-JUL-1999 | 99US-01455276 |
| PR | 27-JUL-1999 | 99US-01455913 |
| PR | 27-JUL-1999 | 99US-01455918 |
| PR | 27-JUL-1999 | 99US-01455919 |
| PR | 28-JUL-1999 | 99US-01459591 |
| PR | 02-AUG-1999 | 99US-0146386  |
| PR | 02-AUG-1999 | 99US-0146388  |
| PR | 02-AUG-1999 | 99US-0146389  |
| PR | 03-AUG-1999 | 99US-0147038  |
| PR | 04-AUG-1999 | 99US-0147204  |
| PR | 05-AUG-1999 | 99US-0147302  |
| PR | 05-AUG-1999 | 99US-0147392  |
| PR | 06-AUG-1999 | 99US-0147260  |
| PR | 06-AUG-1999 | 99US-0147303  |
| PR | 09-AUG-1999 | 99US-0147416  |
| PR | 09-AUG-1999 | 99US-0147493  |
| PR | 09-AUG-1999 | 99US-0147935  |
| PR | 11-AUG-1999 | 99US-0148171  |
| PR | 11-AUG-1999 | 99US-0148177  |
| PR | 11-AUG-1999 | 99US-0148341  |
| PR | 13-AUG-1999 | 99US-0148565  |
| PR | 13-AUG-1999 | 99US-0148684  |
| PR | 16-AUG-1999 | 99US-0149368  |
| PR | 17-AUG-1999 | 99US-0149175  |
| PR | 18-AUG-1999 | 99US-0149246  |
| PR | 20-AUG-1999 | 99US-0149722  |
| PR | 20-AUG-1999 | 99US-0149929  |
| PR | 23-AUG-1999 | 99US-0149920  |
| PR | 23-AUG-1999 | 99US-0149930  |
| PR | 25-AUG-1999 | 99US-0150566  |
| PR | 26-AUG-1999 | 99US-0150884  |
| PR | 27-AUG-1999 | 99US-0151065  |
| PR | 27-AUG-1999 | 99US-0151066  |
| PR | 27-AUG-1999 | 99US-0151080  |
| PR | 30-AUG-1999 | 99US-0151303  |
| PR | 31-AUG-1999 | 99US-0151438  |
| PR | 01-SEP-1999 | 99US-0151930  |
| PR | 01-SEP-1999 | 99US-0152363  |
| PR | 07-SEP-1999 | 99US-0152070  |
| PR | 10-SEP-1999 | 99US-0153758  |
| PR | 13-SEP-1999 | 99US-0153758  |
| PR | 15-SEP-1999 | 99US-0154018  |
| PR | 16-SEP-1999 | 99US-0154039  |
| PR | 20-SEP-1999 | 99US-0154479  |
| PR | 22-SEP-1999 | 99US-0155139  |
| PR | 23-SEP-1999 | 99US-0155486  |
| PR | 24-SEP-1999 | 99US-0155589  |
| PR | 28-SEP-1999 | 99US-0156458  |
| PR | 29-SEP-1999 | 99US-0156596  |
| PR | 04-OCT-1999 | 99US-0157117  |
| PR | 04-OCT-1999 | 99US-0157753  |
| PR | 05-OCT-1999 | 99US-0157865  |
| PR | 06-OCT-1999 | 99US-0158023  |
| PR | 07-OCT-1999 | 99US-0158029  |
| PR | 08-OCT-1999 | 99US-0158359  |
| PR | 12-OCT-1999 | 99US-0158293  |
| PR | 13-OCT-1999 | 99US-0159294  |
| PR | 13-OCT-1999 | 99US-0159295  |
| PR | 14-OCT-1999 | 99US-0159329  |
| PR | 14-OCT-1999 | 99US-0159330  |
| PR | 18-OCT-1999 | 99US-0159331  |
| PR | 18-OCT-1999 | 99US-0159584  |
| PR | 21-OCT-1999 | 99US-0160741  |
| PR | 21-OCT-1999 | 99US-0160767  |
| PR | 21-OCT-1999 | 99US-0160768  |
| PR | 21-OCT-1999 | 99US-0160770  |
| PR | 21-OCT-1999 | 99US-0160814  |
| PR | 21-OCT-1999 | 99US-0160815  |

[illegible]

| Result    | ID       | Accession  | Standard      | DNA | Length (bp) |
|-----------|----------|--|---------------|-----|-------------|
| Result 11 | AAC33292 |  | standard      | DNA | 1239        |
|           | ID       | AAC33292   |               |     |             |
|           | AC       |  |               |     |             |
|           | AC       | AAC33292   |               |     |             |
|           | XX       |  |               |     |             |
|           | XX       |  |               |     |             |
|           | DT       | 17-OCT-2000  | (first entry) |     |             |
|           | XX       |  |               |     |             |
|           | DE       | Arabidopsis thaliana DNA fragment SEQ ID NO: 2492.             |               |     |             |
|           | XX       |  |               |     |             |
|           | KW       | Hybridisation assay; genetic mapping; gene expression control; |               |     |             |
|           | KW       | Protein identification; signal transduction pathway;           |               |     |             |
|           | KW       | metabolic pathway; promoter; termination sequence; ss.         |               |     |             |
|           | XX       |  |               |     |             |
|           | OS       | Arabidopsis thaliana.  |               |     |             |
|           | XX       |  |               |     |             |
|           | EN       | EP1033405-A2.  |               |     |             |
|           | XX       |  |               |     |             |
|           | PD       | 06-SEP-2000.   |               |     |             |
|           | XX       |  |               |     |             |
|           | PF       | 25-FEB-2000; 2000EP-0301439.                                   |               |     |             |
|           | XX       |  |               |     |             |
|           | PR       | 25-FEB-1999; 99US-0121825.                                     |               |     |             |
|           | PR       | 05-MAR-1999; 99US-0123180.                                     |               |     |             |
|           | PR       | 09-MAR-1999; 99US-0123548.                                     |               |     |             |
|           | PR       | 23-MAR-1999; 99US-0125788.                                     |               |     |             |
|           | PR       | 25-MAR-1999; 99US-0126264.                                     |               |     |             |
|           | PR       | 29-MAR-1999; 99US-0127482.                                     |               |     |             |
|           | PR       | 01-APR-1999; 99US-0128234.                                     |               |     |             |
|           | PR       | 06-APR-1999; 99US-0128714.                                     |               |     |             |
|           | PR       | 08-APR-1999; 99US-0129845.                                     |               |     |             |
|           | PR       | 16-APR-1999; 99US-0130077.                                     |               |     |             |
|           | PR       | 19-APR-1999; 99US-0130449.                                     |               |     |             |
|           | PR       | 21-APR-1999; 99US-0130891.                                     |               |     |             |
|           | PR       | 23-APR-1999; 99US-0131449.                                     |               |     |             |
|           | PR       | 28-APR-1999; 99US-0132048.                                     |               |     |             |
|           | PR       | 30-APR-1999; 99US-0132407.                                     |               |     |             |
|           | PR       | 04-MAY-1999; 99US-0132484.                                     |               |     |             |
|           | PR       | 05-MAY-1999; 99US-0132485.                                     |               |     |             |
|           | PR       | 06-MAY-1999; 99US-0132486.                                     |               |     |             |



[illegible]

CC The invention relates to human polynucleotides (AA179941-AA193841) and

CC used to generate reagents and screen small molecule libraries or other

to yield novel antimicrobial compounds. In addition, nucleic acid probes complementary to (I) that are specific for particular species of microorganisms can be used to identify particular microorganism species in clinical specimens, therefore, providing a rapid and dependable method by which to identify the causative agents of a bacterial infection. Also, antibodies generated against proteins translated from mRNA transcribed from proliferation-required sequences can also be used to screen for specific microorganisms that produce such proteins in a species-specific manner. AAH84371 and AAH84670 represent sequencing primers used in the isolation of *E. coli* growth and proliferation related sequence, which are used in an example from the present invention.

Sequence 2022 BP; 495 A; 488 C; 593 G; 446 T; 0 other;

Query Match 6.4%; Score 34.2; DB 22; Length 2022;  
Best Local Similarity 55.5%; Pred. No. 5.4;  
Matches 66; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 334 CTGCCCCCTTCTAAGAACTCGAATGGGTCTCTGCTGCTTACGAGAGTCTCT 393  
DB 312 CTCTCTGTTGTTGAGATACGATCAGCTTGTCTTAAGAGTTAGCGAGGGCT 371  
QY 394 GATTAAACCCAGAGAGATTCTCTCCGAGAGAGAGTGGAGCTAGTACGAGAGA 452  
DB 372 GATTGAAGATGACAAAGTTCTCTCGAACAACGATTGACCATCTAATCGAAGA 430

## RESULT 14

AA565332 standard; cDNA; 663 BP.

AC AAS65332;

DT 13-FEB-2002 (first entry).

DE DNA encoding novel human diagnostic protein #1136.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR P-PSDB; ABG01145.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

PS Claim 1; SEQ ID No 1136; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations in  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.

Sequence 663 BP; 240 A; 95 C; 116 G; 212 T; 0 other;

Query Match 6.4%; Score 33.8; DB 23; Length 663;  
Best Local Similarity 45.6%; Pred. No. 4.3;  
Matches 119; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 28 CAGAGAAATGTCAGTGGTTTGTGCTTAGAGACAAAGTCACTTGTCCAGGCCGAGGTG 87  
DB 335 CATCAAGATTAAAGTATTTGCTTAAAGCTAAATGTCATTAATTTAACTGAG 414  
QY 88 TCACCATAGAAAACCTCTGTATGATCACAACACACACATTCAAATGGTTACCGCA 147  
DB 415 TTATTTGAAAAAAGTTATTTGGCAGAAAAAATTAAGAGATGTATATGTCAAGCA 474  
QY 148 GCACACAGCCCAACACACAGAGGTGTGGAAGCCCAATTTCTGAACTCCAGATATGT 207  
DB 475 GCCACAGAGGGAGAAAGAGAACGTTGGAAGAACCATTAATTTGGAACATGTAGACAC 534  
QY 208 ACTCTTGTACACCTTTTACCTTAAACAAATGCAAGAGAGATTTCTAACCTAAGAAC 267  
DB 535 ATGAGAGGAATTAATCTTAAACCAAGACGAAAGGGGAAAGGTGTTCAATTCTAATAA 594  
QY 268 AGAAGAAAACAAAGAAAGAAC 288  
DB 595 GGAAGAAAAGACTAATATATC 615

## RESULT 15

AAS80842 standard; cDNA; 663 BP.

AC AAS80842;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #16646.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR P-PSDB; ABG16655.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

XX  
PS Claim 1: SEQ ID No 16646; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPRO  
CC at ftp.wipro.int/pub/published\_pcl\_sequences.

XX  
SQ Sequence 663 BP; 240 A; 95 C; 116 G; 212 T; 0 other;

Query Match 6.4%; Score 33.8; DB 23; Length 663;  
Best Local Similarity 45.6%; Pred. No. 4.3;

Matches 119; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 28 CAGAGATGTCAGGTTTGGCTTAGACAAAGTCACCTTGTCCAGGCCAGGTGG 87  
DB 355 CATTCAAGATTAAAGTATTTGGATTAGCTAAATGTCATATTTAAATTTAACTGAG 414  
QY 88 TGACCATAGAAAACCTGTTATGATCACAACACACACATTCMAAATGGTTACGCA 147  
DB 415 TTATTGAAAAAAGTTATTTGGGAAAGAAAAAATTAAAGAGATGTAATATGTCAGACA 474  
QY 148 GCACACAGCCCAACACACAGAGGTGTGGAAGCCAAATTTCTGACGTCAGGATATGT 207  
DB 475 GCCACAGAGGGAAGAGAACAGTGTGAAGAACCATATATGATTGGAAACATGTAGAAC 534  
QY 208 ACTCTTCGTACACCTTTACTTAATAAATGACAGAGAGATTTCTACACCTTAAGAAC 267  
DB 535 ATGAGGAAATTTACTTAAACCAAGAAAGCAAGGGAAGAGTGTCTATTTCTAATAAA 594  
QY 268 AGAAGAAAACAAGAAAGAAC 288  
DB 595 GGAAGAAAAGAGTAATAATC 615

Search completed: October 19, 2002, 05:54:31  
Job time : 225 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 19, 2002, 05:44:31 ; Search time 44 Seconds  
(without alignments)  
2969,932 Million cell updates/sec

Title: AI327498

Sequence: 1 AATTCGGCAGTACGACACCC.....CATTAATTCCTGTCATCAT 532

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/prodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/prodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/prodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/prodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/prodata/2/ina/PCUTUS.COMB.seq:\*  
6: /cgn2\_6/prodata/2/ina/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description       |
|------------|-------|-------------|--------|-------|-------------------|
| C 1        | 38.8  | 7.3         | 7218   | 1     | US-08-232-463-14  |
| 2          | 33.8  | 6.4         | 5163   | 3     | US-08-700-651-1   |
| 3          | 33.8  | 6.4         | 5163   | 3     | US-08-928-361B-4  |
| 4          | 33.8  | 6.4         | 5318   | 3     | US-08-700-651-2   |
| 5          | 33.8  | 6.4         | 5318   | 3     | US-08-928-361B-3  |
| 6          | 33.6  | 6.3         | 7286   | 3     | US-09-331-581-3   |
| 7          | 33.6  | 6.3         | 7938   | 3     | US-09-331-581-14  |
| 8          | 33.2  | 6.2         | 289    | 4     | US-09-007-005-17  |
| 9          | 33.2  | 6.2         | 289    | 4     | US-09-244-796-17  |
| 10         | 32.6  | 6.1         | 5014   | 4     | US-09-381-862-3   |
| 11         | 32    | 6.0         | 2139   | 3     | US-09-026-343-1   |
| 12         | 32    | 6.0         | 6924   | 1     | US-08-915-973-2   |
| 13         | 32    | 6.0         | 6924   | 2     | US-08-448-164-2   |
| 14         | 31.8  | 6.0         | 1099   | 4     | US-08-944-604-19  |
| 15         | 31.8  | 6.0         | 1262   | 4     | US-08-944-604-17  |
| 16         | 31.8  | 6.0         | 1331   | 4     | US-08-944-604-17  |
| 17         | 31.8  | 6.0         | 4621   | 1     | US-08-770-761A-1  |
| 18         | 31.8  | 6.0         | 5511   | 3     | US-08-928-361B-2  |
| 19         | 31.8  | 6.0         | 7334   | 3     | US-08-928-361B-1  |
| 20         | 31.8  | 6.0         | 19124  | 2     | US-08-487-828B-13 |
| 21         | 31.2  | 5.9         | 198    | 5     | PCT-US95-10668-1  |
| 22         | 31.2  | 5.9         | 198    | 5     | PCT-US95-10668-2  |
| 23         | 30.8  | 5.8         | 2657   | 2     | US-08-974-565C-8  |
| 24         | 30.8  | 5.8         | 2657   | 2     | US-09-255-748-8   |
| 25         | 30.6  | 5.8         | 1296   | 4     | PCT-US91-00909-5  |
| 26         | 30.6  | 5.8         | 5024   | 4     | US-09-307-143-1   |
| 27         | 30.6  | 5.8         | 246240 | 2     | US-08-724-394A-20 |

|      |      |     |        |   |                   |                   |
|------|------|-----|--------|---|-------------------|-------------------|
| C 28 | 30.6 | 5.8 | 246240 | 2 | US-08-724-394A-21 | Sequence 21, Appl |
| C 29 | 30.6 | 5.8 | 246240 | 2 | US-08-724-394A-22 | Sequence 22, Appl |
| 30   | 30.2 | 5.7 | 1066   | 1 | US-08-157-101A-4  | Sequence 4, Appl  |
| 31   | 30.2 | 5.7 | 1398   | 2 | US-08-896-320-2   | Sequence 2, Appl  |
| 32   | 30.2 | 5.7 | 6601   | 2 | US-09-356-952-10  | Sequence 10, Appl |
| 33   | 29.8 | 5.6 | 4808   | 3 | US-08-231-193A-10 | Sequence 10, Appl |
| 34   | 29.8 | 5.6 | 4808   | 3 | US-08-486-273A-10 | Sequence 10, Appl |
| 35   | 29.8 | 5.6 | 4808   | 3 | US-08-940-035A-10 | Sequence 10, Appl |
| 36   | 29.8 | 5.6 | 4808   | 3 | US-08-940-035A-10 | Sequence 10, Appl |
| 37   | 29.8 | 5.6 | 4858   | 3 | PCT-US95-10668-3  | Sequence 9, Appl  |
| 38   | 29.6 | 5.6 | 198    | 5 | PCT-US95-10668-4  | Sequence 9, Appl  |
| 39   | 29.6 | 5.6 | 1107   | 2 | US-08-991-300-1   | Sequence 4, Appl  |
| 40   | 29.6 | 5.6 | 1219   | 4 | US-09-230-380-10  | Sequence 10, Appl |
| C 41 | 29.6 | 5.6 | 1325   | 1 | US-08-306-691B-51 | Sequence 51, Appl |
| C 42 | 29.6 | 5.6 | 2151   | 4 | US-09-626-589-4   | Sequence 4, Appl  |
| C 43 | 29.6 | 5.6 | 2284   | 2 | US-08-467-822-28  | Sequence 28, Appl |
| C 44 | 29.6 | 5.6 | 2284   | 2 | US-08-432-637-28  | Sequence 28, Appl |
| 45   | 29.6 | 5.6 | 2284   | 4 | US-08-432-637-28  | Sequence 28, Appl |

## ALIGNMENTS

RESULT 1  
US-08-232-463-14/C

/ Sequence 14, Application US/08232463

/ Patent No. 5670367

/ GENERAL INFORMATION:

/ APPLICANT: DORNER, F.

/ APPLICANT: FALKNER, F. G.

/ TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

/ NUMBER OF SEQUENCES: 52

/ CORRESPONDENCE ADDRESS:

/ ADDRESSEE: Foley & Lardner

/ STREET: 1800 Diagonal Road, Suite 500

/ CITY: Alexandria

/ STATE: VA

/ COUNTRY: USA

/ ZIP: 22313-0299

/ COMPUTER READABLE FORM:

/ MEDIUM TYPE: Floppy disk

/ COMPUTER: IBM PC compatible

/ OPERATING SYSTEM: PC-DOS/MS-DOS

/ SOFTWARE: Patentin Release #1.0, Version #1.25

/ CURRENT APPLICATION DATA:

/ APPLICATION NUMBER: US/08/232,463

/ FILING DATE:

/ CLASSIFICATION: 435

/ PRIOR APPLICATION DATA:

/ APPLICATION NUMBER: US/07/935,313

/ FILING DATE:

/ ATTORNEY/AGENT INFORMATION:

/ NAME: BENT, Stephen A.

/ REGISTRATION NUMBER: 29,768

/ REFERENCE/DOCKET NUMBER: 30472/114

/ TELECOMMUNICATION INFORMATION:

/ TELEPHONE: (703)836-9300

/ TELEFAX: (703)683-4109

/ INFORMATION FOR SEQ ID NO: 14:

/ SEQUENCE CHARACTERISTICS:

/ LENGTH: 7218 base pairs

/ TYPE: nucleic acid

/ STRANDEDNESS: single

/ TOPOLOGY: linear

/ IMMEDIATE SOURCE:

/ CLONE: pTZ9pC-F15

/ US-08-232-463-14

Query Match

7.3%; Score 38.8; DB 1; Length 7218;

Best Local Similarity 3.4%; Pred. No. 0.083; Matches 13; Conservative 208; Mismatches 165; Indels 0; Gaps 0;

```

QY 93 ATGAGAACTCTGTTATGATCACAACAACACCATTCATAAATGTTACCGAGGACA 152
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1451 ATGAGAAATTTGTCACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1392
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 153 CAGCCACAACAACGAGGTTGGAAGCCATTTCTGAAGCTCAGATATGACTCT 212
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1391 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1332
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 213 TCGACACCTTACTTAAATAATGACAGAGAGATTTCTAACAATAAGAAACAGAG 272
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1331 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1272
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 273 AAAACAAGAAAGAACGGGCTTTTGAGCGGGAATGACACATGCGCCAGGGCTCAC 332
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1271 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1212
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 333 CCGGCCCTTGTCTAAGGAATCTGGAATGGGTCTCTGCTGCTCAGGAGTCTC 392
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1211 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1152
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 393 TGATTAAACCCAGAGGAGTTCTCTCCGAGAGAGAGAGTGAAGTACGTTAGCCAGAGA 452
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1151 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1092
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 453 GAAAGGATGTGGTGAGAGGAGAAA 478
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1091 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1066
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 2  
US-08-700-651-1  
Sequence 1, Application US/08700651B

```

; GENERAL INFORMATION:
; PATENT NO. 6015882
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; TITLE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; EARLIER FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 5163
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
US-08-700-651-1

```

Query Match 6.4%; Score 33.8; DB 3; Length 5163;  
Best Local Similarity 49.7%; Pred. No. 2.1;  
Matches 86; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

```

QY 115 ACAACAACAACACATTCATAAATGTTACCGACACACAGCCCAACAACACAGAGGTG 174
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 674 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 733
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 175 TGAAGCCATTTCTGAAGCTCCAGATATGTAATCTTCTGACACCTTACCTAAAAA 234
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 734 ACGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 793
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 235 ACATGCAAGAGAGATTTTCAACCTAAGAAACAAGAAAGAAAGAAAGAGAAC 287
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 794 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 846
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 3  
US-08-928-361B-4  
Sequence 4, Application US/08928361B

```

; GENERAL INFORMATION:
; PATENT NO. 6071518
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: VERNY, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1678
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5163 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-928-361B-4

```

Query Match 6.4%; Score 33.8; DB 3; Length 5163;  
Best Local Similarity 49.7%; Pred. No. 2.1;  
Matches 86; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

```

QY 115 ACAACAACAACACATTCATAAATGTTACCGACACACAGCCCAACAACACAGAGGTG 174
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 673 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 732
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 175 TGAAGCCATTTCTGAAGCTCCAGATATGTAATCTTCTGACACCTTACCTAAAAA 234
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 733 ACGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 792
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 235 ACATGCAAGAGAGATTTTCAACCTAAGAAACAAGAAAGAAAGAAAGAGAAC 287
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 793 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 845
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 4  
US-08-700-651-2  
Sequence 2, Application US/08700651B

```

; GENERAL INFORMATION:
; PATENT NO. 6015882
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI

```

```

; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700.651B
; EARLIER FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415.751
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5318
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
US-08-700-651-2

```

```

Query Match
Best Local Similarity 49.7%; Pred. No. 2.1;
Matches 86; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

```

```

OY 115 ACAACACACACACACATTCATAAATGGTTACCGCACACACACCCACACACAGGAGGTGG 174
Db 674 ACAACACACACACACACACACACACACACACACACACACACACACACACACACACAC 733
OY 175 TGGAAAGCCAAATTTCTGACGCTCCAGATATGTACTCTTGTACCACTTTACTTAATAA 234
Db 734 ACGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 793
OY 235 ACATGCAGAAAGAGATTTCTACACACTTAAGAACAGAAAGAAAGAAAGAGAAC 287
Db 794 ACAACCAACTACTACACACACACACACACACACACACACACACACACACACACAC 846

```

# RESULT 5

```

US-08-928-361B-3
; Sequence 3, Application US/08928361B
; Patent No. 6071518

```

## GENERAL INFORMATION:

```

; APPLICANT: Peteresen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIRKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840

```

## COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 12-SEP-1997
; APPLICATION NUMBER: US/08/928.361B
; CLASSIFICATION:

```

## PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: VERNY, HANA
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO. 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5318 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-928-361B-3

```

```

Query Match
Best Local Similarity 49.7%; Pred. No. 2.1;
Matches 86; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

```

```

OY 115 ACAACACACACACACATTCATAAATGGTTACCGCACACACACCCACACACAGGAGGTGG 174
Db 673 ACAACACACACACACACACACACACACACACACACACACACACACACACACACACAC 732
OY 175 TGGAAAGCCAAATTTCTGACGCTCCAGATATGTACTCTTGTACCACTTTACTTAATAA 234
Db 733 ACGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 792
OY 235 ACATGCAGAAAGAGATTTCTACACACTTAAGAACAGAAAGAAAGAAAGAGAAC 287
Db 793 ACAACCAACTACTACACACACACACACACACACACACACACACACACACACACAC 845

```

# RESULT 6

```

US-09-331-581-3
; Sequence 3, Application US/09331581
; Patent No. 6130070

```

## GENERAL INFORMATION:

```

; APPLICANT: TOHDA, Hideki
; APPLICANT: HAMA, Yuko
; APPLICANT: KUMAGAI, Hiromichi
; TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
; TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE. EXPRESSION VECTORS
; FILE REFERENCE: 0059-1142-OPCT
; CURRENT APPLICATION NUMBER: US/09/331,581
; EARLIER FILING DATE: 1999-06-30
; EARLIER APPLICATION NUMBER: PCT/JP98/04929
; EARLIER FILING DATE: 1998-10-30
; EARLIER APPLICATION NUMBER: JP 9-314608
; EARLIER FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 3
; LENGTH: 7286

```

```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:DNA
US-09-331-581-3

```

```

Query Match
Best Local Similarity 6.3%; Score 33.6; DB 3; Length 7286;
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

```

```

OY 230 AAAAAACATGCAGAAAGAGATTTCTCAACCTTAAGAAAGAAAGAAAGAAAGAAAGAAAG 289
Db 4618 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4677
OY 290 GCACCTTTGGACCGGAAATGACACACTGCGCCCGAG 325
Db 4678 GTACCTTTGTGAGCGGAAAGAACACGCGGATCCAG 4713

```

# RESULT 7

```

US-09-331-581-14
; Sequence 14, Application US/09331581
; Patent No. 6130070

```

```

; GENERAL INFORMATION:
; APPLICANT: TOHDA, Hideki
; APPLICANT: HAMA, Yuko
; APPLICANT: KUMAGAI, Hiromichi
; TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE

```



TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS  
FILE OF INVENTION: HAVING THE SAME, AND USE THEREOF  
CURRENT FILING DATE: 1999-06-30  
CURRENT FILING DATE: 1999-06-30  
EARLIER FILING DATE: 1998-10-30  
EARLIER FILING DATE: 1997-10-31  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 14  
LENGTH: 7938  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:DNA  
US-09-331-581-14

Query Match 6.3%; Score 33.6; DB 3; Length 7938;  
Best Local Similarity 59.4%; Pred. No. 2.9;  
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 230 AAAAAATGACGAGAGATCTCTACACCTAAGAAACAGAGAAACAGAGACCG 289  
Db 5270 AA 5329  
Qy 290 GCACCTTTGAGCGGGAATGACACACTGCGCCAG 325  
Db 5330 GTACCTTGTGAGCGGGAATGACACACTGCGCCAG 5365

## RESULT 8

US-09-007-005-17  
Sequence 17, Application US/09007005B  
Patent No. 6258558  
GENERAL INFORMATION:  
APPLICANT: Szostak, Jack W.  
APPLICANT: Roberts, Richard W.  
APPLICANT: Liu, Rihue  
TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN  
FILE REFERENCE: 00786/350003  
CURRENT FILING DATE: 1998-01-14  
EARLIER FILING DATE: 1997-01-27  
EARLIER FILING DATE: 1997-01-27  
EARLIER FILING DATE: 1997-11-06  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 17  
LENGTH: 289  
TYPE: RNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Translation template  
NAME/KEY: misc-feature  
LOCATION: (1)..(289)  
OTHER INFORMATION: n = A,T,C or G  
US-09-007-005-17

Query Match 6.2%; Score 33.2; DB 4; Length 289;  
Best Local Similarity 9.3%; Pred. No. 0.98;  
Matches 25; Conservative 100; Mismatches 143; Indels 0; Gaps 0;

Qy 239 GCAGAGAGATCTCTACACCTAAGAAACAGAGAAACAGAGACCGCCTTTG 298  
Db 6 GRARARARURURARURURURARURURURARURURURARURURURARUR 65  
Qy 299 GAGCGGAATGACACACTGCGCCCGGCTACCCCTGCTGTAAGGAATCCTG 358

Db 66 NRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNR 125  
Qy 359 AATGGGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 418  
Db 126 NRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNR 185  
Qy 419 CGAA 478  
Db 186 NRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNR 245  
Qy 479 GACTTGAGCGAGCTGCAAGAAATCAA 506  
Db 246 CRURURGRGRGRURUAAAAAAAAAAAA 273

RESULT 9  
US-09-244-796-17  
Sequence 17, Application US/09244796  
Patent No. 6281344  
GENERAL INFORMATION:  
APPLICANT: Szostak, Jack W.  
APPLICANT: Roberts, Richard W.  
APPLICANT: Liu, Rihue  
TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN  
FILE REFERENCE: 00786/350007  
CURRENT FILING DATE: 1999-02-05  
EARLIER FILING DATE: 1997-01-27  
EARLIER FILING DATE: 1997-01-27  
EARLIER FILING DATE: 1997-11-06  
EARLIER FILING DATE: 09/007, 005  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 17  
LENGTH: 289  
TYPE: RNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Translation template  
NAME/KEY: misc-feature  
LOCATION: (1)..(289)  
OTHER INFORMATION: n = A,T,C or G  
US-09-244-796-17

Query Match 6.2%; Score 33.2; DB 4; Length 289;  
Best Local Similarity 9.3%; Pred. No. 0.98;  
Matches 25; Conservative 100; Mismatches 143; Indels 0; Gaps 0;

Qy 239 GCAGAGAGATCTCTACACCTAAGAAACAGAGAAACAGAGACCGCCTTTG 298  
Db 6 GRARARARURURARURURURARURURURARURURURARURURURARUR 65  
Qy 299 GAGCGGAATGACACACTGCGCCCGGCTACCCCTGCTGTAAGGAATCCTG 358  
Db 66 NRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNR 125  
Qy 359 AATGGGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 418  
Db 126 NRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNR 185  
Qy 419 CGAA 478  
Db 186 NRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNR 245  
Qy 479 GACTTGAGCGAGCTGCAAGAAATCAA 506  
Db 246 CRURURGRGRGRURUAAAAAAAAAAAA 273

```

? APPLICANT: CONAWAY, JOAN W.
? APPLICANT: CONAWAY, RONALD C.
? TITLE OF INVENTION: ELI2, A New Member of an ELL Family of
? TITLE OF INVENTION: RNA Polymerase II Elongation Factors
? NUMBER OF SEQUENCES: 34
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
? STREET: 1100 NEW YORK AVENUE, SUITE 600
? CITY: WASHINGTON
? STATE: D.C.
? COUNTRY: USA
? ZIP: 20005-3934
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentln Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/026,343
? FILING DATE: Herewith
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 60/036,447
? FILING DATE: 19-FEB-1997
? ATTORNEY/AGENT INFORMATION:
? NAME: GOLDSTEIN, JORGE A.
? REGISTRATION NUMBER: 29,021
? REFERENCE/DOCKET NUMBER: 1488.0880001
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 371-2600
? TELEFAX: (202) 372-2540
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2139 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: both
? MOLECULE TYPE: CDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 94..2013
? US-09-026-343-1

Query Match          6.0%; Score 32; DB 3; Length 2139;
Best Local Similarity 58.3%; Pred.No. 4.9;
Matches    56; Conservative      0; Mismatches   40; Indels     0; Gaps      0

QY 248 GATCTACACCCATAAAGAAACAAGAAGAAAACAAGAAGACCAGCGACTTTTGGAGCCGGA 307
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1100 GAATTCTCCTCGAGCAAGACAGATGCCAACACATAATGTGCATTGAATCCACCA 1159
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 308 ATGACACACTGGCCCCCAGGCGCTCACCCCTGCCCTTG 343
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1160 GTGAAAAATCGCTGCAGCGCTCCACACAGCCGCCCTG 1195
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
US-08-015-973-2
? Sequence 2, Application US/08015973
? Patent No. 5604094
? GENERAL INFORMATION:
? APPLICANT: Schlessinger, Joseph
? TITLE OF INVENTION: A NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
? TITLE OF INVENTION: PHOSPHATASE-BETA
? NUMBER OF SEQUENCES: 9
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: PENNIE & EDMONDS
? STREET: 1155 Avenue of the Americas
? CITY: New York
? STATE: New York
? COUNTRY: U.S.A.
? ZIP: 10036-2711
? COMPUTER READABLE FORM:
```

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/015,973  
FILING DATE: 10-FEB-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MISTOCK, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7683-021  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)790-9090  
TELEFAX: (212)869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6924 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..6924  
US-08-015-973-2

Query Match  
Best Local Similarity 6.0%; Score 32; DB 1; Length 6924;  
Matches 47; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 16 CACCCCCCTGCAGAGATGTCAGTGTGCTTAGAGACAAGTCACTTGTCCA 75  
DB 710 CTCCTCCCTGCAGACAGACAGTGTGATGATGTTTAAAGATACAGTTAGCATCTCTG 769

QY 76 AGGCCAGGTGG 87  
DB 770 AAGCCAGTTGG 781

RESULT 13  
US-08-448-164-2  
Sequence 2, Application US/08448164  
Patent No. 5925536  
GENERAL INFORMATION:  
APPLICANT: Schlensing, Joseph  
TITLE OF INVENTION: A NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/448,164  
FILING DATE: 24-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/015,973  
FILING DATE: 10-FEB-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: MISTOCK, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7683-021

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)790-9090  
TELEFAX: (212)869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6924 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..6924  
US-08-448-164-2

Query Match  
Best Local Similarity 6.0%; Score 32; DB 2; Length 6924;  
Matches 47; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 16 CACCCCCCTGCAGAGATGTCAGTGTGCTTAGAGACAAGTCACTTGTCCA 75  
DB 710 CTCCTCCCTGCAGACAGACAGTGTGATGATGTTTAAAGATACAGTTAGCATCTCTG 769

QY 76 AGGCCAGGTGG 87  
DB 770 AAGCCAGTTGG 781

RESULT 14  
US-08-944-604-19  
Sequence 19, Application US/08944604  
Patent No. 6218131  
GENERAL INFORMATION:  
APPLICANT: KERSEY, SUSAN  
APPLICANT: OBAR, ROBERT  
TITLE OF INVENTION: MATERIALS AND METHODS FOR DETECTION OF  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Testa, Hurwitz & Thibault  
STREET: 125 High St.  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/944,604  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MEYERS, THOMAS C  
REGISTRATION NUMBER: 36,989  
REFERENCE/DOCKET NUMBER: MTP-021 (8395/24)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7000  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1099 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 61..948

OTHER INFORMATION: /product="BC-8 (Isoform B)"  
US-08-944-604-19

Query Match 6.0%; Score 31.8; DB 4; Length 1099;  
Best Local Similarity 49.7%; Pred. No. 4.3;  
Matches 81; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 327 GCTCACCCCTGCCCCCTTGTCTAAGAAATCCTGGAAATGTGGCTCTCTGCTTCCCTTCAGCG 386  
DB 470 GCTCAAAATCCTCAGGTGTACATGACATCAAGATTGGGAACAAGCCGGCTGCGCATCC 529  
QY 387 AGTCTGTATTAAACCCAGAGGAGTTCTCTCCGAGAGAAGAGTGGAGCTAGCTTACGCC 446  
DB 530 AGATGCTCTGCTGCTTGTGATGTGTGCTGCCATGACACAGAGAAATTTCCGCTGTGCA 589  
QY 447 AGAAGAGAAAGGAGTGTGGTGTGAGAGGAAAGACCTTCGACGG 489  
DB 590 CTCATGAAAGGGCTTGTGGCTTAAAGGAAGACGCTTCACCG 632

RESULT 15

US-08-989-386-4  
; Sequence 4, Application US/08989386  
; Patent No. 5989860

GENERAL INFORMATION:

APPLICANT: Bandman, Olga  
APPLICANT: Hillman, Jennifer L.

APPLICANT: Guegler, Karl J.

APPLICANT: Corley, Neil C.

APPLICANT: Shah, Purvi

TITLE OF INVENTION: HUMAN ISOMERASE HOMOLOGS

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: PastSEO for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/989,386

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0443 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555

TELEFAX: 650-845-4166

TELEX:

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1262 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: TMLRDT01

CLONE: 292808

US-08-989-386-4

Query Match 6.0%; Score 31.8; DB 2; Length 1262;  
Best Local Similarity 49.7%; Pred. No. 4.6;  
Matches 81; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 327 GCTCACCCCTGCCCCCTTGTCTAAGAAATCCTGGAAATGTGGCTCTCTGCTTCCCTTCAGCG 386  
DB 468 GCTCAAAATCCTCAGGTGTACATGACATCAAGATTGGGAACAAGCCGGCTGCGCATCC 527  
QY 387 AGTCTGTATTAAACCCAGAGGAGTTCTCTCCGAGAGAAGAGTGGAGCTAGCTTACGCC 446  
DB 528 AGATGCTCTGCTGCTTGTGATGTGTGCTGCCATGACACAGAGAAATTTCCGCTGTGCA 587  
QY 447 AGAAGAGAAAGGAGTGTGGTGTGAGAGGAAAGACCTTCGACGG 489  
DB 588 CTCATGAAAGGGCTTGTGGCTTAAAGGAAGACGCTTCACCG 630

Search completed: October 19, 2002, 06:54:48  
Job time : 65 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 19, 2002, 04:58:21 ; Search time 1637 Seconds

(Without alignments)  
4386.304 Million cell updates/sec

Title: AI327498

Sequence: 1 AATTCGCACCTAGACACCC.....CATTAATTCCTGCAATCAT 532

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estinu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estli:\*  
10: gb\_estlz:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inu:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1          | 532   | 100.0       | 532    | 9     | AI327498    |
| 2          | 516.4 | 97.1        | 704    | 10    | BG865668    |
| 3          | 516.4 | 97.1        | 742    | 10    | BG871142    |
| 4          | 516.4 | 97.1        | 757    | 10    | BG866610    |
| 5          | 516.4 | 97.1        | 1361   | 11    | BC018582    |
| 6          | 514.8 | 96.8        | 874    | 10    | BG974340    |
| 7          | 513.8 | 96.6        | 841    | 10    | BG867642    |
| 8          | 511.6 | 96.2        | 806    | 10    | BF299645    |
| 9          | 505.4 | 95.0        | 822    | 10    | BF300481    |
| 10         | 504.4 | 94.8        | 740    | 10    | BG866897    |
| 11         | 504.4 | 94.8        | 849    | 10    | BF537367    |
| 12         | 504.4 | 94.8        | 891    | 10    | BF535022    |
| 13         | 503.8 | 94.7        | 743    | 10    | BG873314    |
| 14         | 502.8 | 94.5        | 988    | 10    | BF300873    |
| 15         | 496.4 | 93.3        | 741    | 10    | BG871007    |
| 16         | 494.8 | 93.0        | 708    | 10    | BG868576    |
| 17         | 493.4 | 92.7        | 827    | 10    | BG865197    |

|    |       |      |     |    |          |
|----|-------|------|-----|----|----------|
| 18 | 487.4 | 91.6 | 675 | 9  | AV249432 |
| 19 | 487.4 | 91.6 | 680 | 10 | BF300518 |
| 20 | 485.4 | 91.2 | 860 | 10 | BG866108 |
| 21 | 483.4 | 90.9 | 682 | 10 | BF535168 |
| 22 | 482.4 | 90.7 | 729 | 10 | BG870130 |
| 23 | 480.8 | 90.4 | 631 | 10 | BF301693 |
| 24 | 480.4 | 90.3 | 727 | 10 | BF539392 |
| 25 | 479   | 90.0 | 791 | 10 | BG869048 |
| 26 | 478   | 89.8 | 810 | 10 | BF301149 |
| 27 | 474.4 | 89.2 | 730 | 10 | BG871270 |
| 28 | 470.2 | 88.4 | 740 | 10 | BG869019 |
| 29 | 469.4 | 88.2 | 678 | 10 | BG869045 |
| 30 | 468.6 | 88.1 | 974 | 10 | BG870966 |
| 31 | 465.4 | 87.5 | 955 | 10 | BF538505 |
| 32 | 464.4 | 87.3 | 718 | 10 | BF540048 |
| 33 | 462.8 | 87.0 | 826 | 10 | BG865309 |
| 34 | 462   | 86.8 | 851 | 10 | BG872096 |
| 35 | 461.4 | 86.7 | 799 | 10 | BG871270 |
| 36 | 459.8 | 86.4 | 679 | 10 | BG871915 |
| 37 | 459.6 | 86.4 | 703 | 10 | BF534356 |
| 38 | 459.4 | 86.4 | 692 | 10 | BG868162 |
| 39 | 456.8 | 85.9 | 834 | 10 | BG870992 |
| 40 | 456.6 | 85.8 | 672 | 10 | BF539409 |
| 41 | 456.4 | 85.8 | 665 | 10 | BF538109 |
| 42 | 453.8 | 85.3 | 734 | 10 | BF534301 |
| 43 | 452.4 | 85.0 | 692 | 10 | BF536890 |
| 44 | 449.8 | 84.5 | 671 | 10 | BF534781 |
| 45 | 448.2 | 84.2 | 670 | 10 | BF300142 |

## ALIGNMENTS

RESULT 1  
LOCUS AI327498  
DEFINITION mb20a07.y1 Soares mouse p3NM19.5 Mus musculus CDNA clone  
IMAGE:329940 5' similar to SW:CAH6\_SHEEP P08060 CARBONIC ANHYDRASE  
VI: mRNA sequence.

ACCESSION AI327498.1 GI:4061927  
VERSION AI327498  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 532)  
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Maria M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu

This clone is available royalty-free through LNC; contact the  
IMAGE Consortium (info@image.lln.gov) for further information.  
MGI:211340

This read is a RESSEQUENCE of a previously sequenced mouse clone  
correct orientation)  
Seq primer: -40RP from Gibco  
High quality sequence stop: 476.  
Location/Qualifiers

FEATURES  
Source

1. 532  
/db\_xref="taxon:10090"





mRNA sequence.  
 ACCESSION BG871142  
 VERSION BG871142.1 GI:14221682  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NIH-MGC http://mgc.nci.nih.gov/  
 1 (bases 1 to 742)  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
 Plate: LLAM10840 row: e column: 13  
 High quality sequence stop: 741.  
 Location/Qualifiers  
 1..742  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone\_image="4921716"  
 /clone\_lib="NCI\_CGAP\_SG2"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: salivary gland; Vector: pCMV-Sport6; Site: 1; NotI: Site 2; SalI: Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.3 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP Library."  
 BASE COUNT 231 a 199 c 166 g 146 t  
 ORIGIN  
 Query Match 97.1%; Score 516.4; DB 10; Length 742;  
 Best Local Similarity 99.8%; Pred. No. 3e-105;  
 Matches 517; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 15 ACACCCCCCGCAGAGATGTCAGTGTGTCTTACGACAAAGTCACTTGTCC 74  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 107 ACACCCCCCGCAGAGATGTCAGTGTGTCTTACGACAAAGTCACTTGTCC 166  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 75 AAGCCCGAGGTGTGACATAGAAACTCGTTATGATCAACAACACCACTTCAA 134  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 167 AAGCCCGAGGTGTGACATAGAAACTCGTTATGATCAACAACACCACTTCAA 226  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 135 AATGTTACCGCAGCACAAGCCCAACAACACAGGTGTGGAAGCCATTCTGAAC 194  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 227 AATGTTACCGCAGCACAAGCCCAACAACACAGGTGTGGAAGCCATTCTGAAC 286  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 195 GTCCAGATATGATCTTCTGATACACCTTCTAATAAATATGCAAGAGATTCTA 254  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 287 GTCCAGATATGATCTTCTGATACACCTTCTAATAAATATGCAAGAGATTCTA 346  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 255 CAACCTTAAGAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 314  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 347 CAACCTTAAGAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 406  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 315 ACTGGCCCGAGGCTCACCTGCTCTTGTCTTAAGAAATCTGGAATGTGGTCTCTTG 374  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 407 ACTGGCCCGAGGCTCACCTGCTCTTGTCTTAAGAAATCTGGAATGTGGTCTCTTG 466  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 375 GTGCCCTTACGAGTCTCTGATTAACCGAGAGGAGTTCTCTCGAGAGAAAGTGA 434  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 467 GTGCCCTTACGAGTCTCTGATTAACCGAGAGGAGTTCTCTCGAGAGAAAGTGA 526  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 435 GCTAGCTTACCGAAGAAAGAGATGTGGGTGAGAGGAAAGACTTCAGCGAGCTG 494  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 527 GCTAGCTTACCGAAGAAAGAGATGTGGGTGAGAGGAAAGACTTCAGCGAGCTG 586

RESULT 4  
 LOCUS BG866610  
 DEFINITION 602785586F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:4911884 5',  
 mRNA sequence.  
 ACCESSION BG866610  
 VERSION BG866610.1 GI:14217150  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NIH-MGC http://mgc.nci.nih.gov/  
 1 (bases 1 to 757)  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
 Plate: LLAM10814 row: k column: 21  
 High quality sequence stop: 718.  
 Location/Qualifiers  
 1..757  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone\_image="4911884"  
 /clone\_lib="NCI\_CGAP\_SG2"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: salivary gland; Vector: pCMV-Sport6; Site: 1; NotI: Site 2; SalI: Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.3 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP Library."  
 BASE COUNT 236 a 193 c 164 g 164 t  
 ORIGIN  
 Query Match 97.1%; Score 516.4; DB 10; Length 757;  
 Best Local Similarity 99.8%; Pred. No. 3e-105;  
 Matches 517; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 15 ACACCCCCCGCAGAGATGTCAGTGTGTCTTACGACAAAGTCACTTGTCC 74  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 52 ACACCCCCCGCAGAGATGTCAGTGTGTCTTACGACAAAGTCACTTGTCC 111  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 75 AAGCCCGAGGTGTGACATAGAAACTCTGTTATGATCAACAACACCACTTCAA 134  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 112 AAGCCCGAGGTGTGACATAGAAACTCTGTTATGATCAACAACACCACTTCAA 171  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 135 AATGTTACCGCAGCACAAGCCCAACAACACAGGTGTGGAAGCCATTCTGAAC 194  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 172 AATGTTACCGCAGCACAAGCCCAACAACACAGGTGTGGAAGCCATTCTGAAC 231  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 195 GTCCAGATATGATCTTCTGATACACCTTCTAATAAATATGCAAGAGATTCTA 254  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 232 GTCCAGATATGATCTTCTGATACACCTTCTAATAAATATGCAAGAGATTCTA 291  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 255 CAACCTTAAGAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 314  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 292 CAACCTTAAGAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 351  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 315 ACTGGCCCGAGGCTCACCTGCTCTTGTCTAAGAAATCTGGAATGTGGTCTCTTG 374

```

|||||
Db 352 ACTGCCCCGAGGCTACCCCTGCCCCCTTGTCTAAGAAATCCCTGGATGTGGTCTCCCT 411
Oy 375 CTGCTTCAGCAGCTCTGTGATTAAACAGAGAGGATTCCTCCGAGAGAAGTGA 434
Db 412 CTGCTTCAGCAGCTCTGTGATTAAACAGAGAGGATTCCTCCGAGAGAAGTGA 471
Oy 435 GCTACGTTAGCCAGAGAAGAAAGGATGTGGTGAGAGGAAAGACTTCGACGACGTG 494
Db 472 GCTACGTTAGCCAGAGAAGAAAGGATGTGGTGAGAGGAAAGACTTCGACGACGTG 531
Oy 495 CAAGAATCAAGCCTTCATTAATTCCTGTCATCAT 532
Db 532 CAAGAATCAAGCCTTCATTAATTCCTGTCATCAT 569

RESULT 5
LOCUS BC018582 1361 bp mRNA linear HTC 03-DEC-2001
DEFINITION Mus musculus, clone IMAGE:4163477, mRNA.
ACCESSION BC018582
VERSION BC018582.1 GI:11235764
KEYWORDS HTC.
SOURCE house mouse.
ORGANISM Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1361)
Strausberg, R.
REFERENCE Direct Submission
Submitted (03-DEC-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: villalondcm.tmc.edu.
Villalon, D.K., Luna, R.A., Hale, S.M., Huliyk, S., Lu, X., Garcia,
A.M., Hollway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 30 Row: m Column: 7
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, similarity but not identity to protein
This clone has the following problem: no 5' EST match.
Location/Qualifiers
1. 1361
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:4163477"
/tissue_type="Liver, normal, 5 month old male mouse."
/clone_lib="NCI_CGAP_L19"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
BASE COUNT 401 a 330 c 328 g 302 t
ORIGIN
Query Match 97.1%; Score 516.4; DB 11; Length 1361;
Best Local Similarity 99.8%; Pred. No. 3,1e-105;
Matches 517; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 15 ACACCCCTGCAACAGAAATGTCCAGTGTGTTGCTTAGAGACAAGTCACTTTGTCC 74

```

```

|||||
Db 661 ACACCCCTGCAACAGAAATGTCCAGTGTGTTGCTTAGAGACAAGTCACTTTGTCC 720
Oy 75 AAGGCCAGGTGTGACCATAGAAACCTGTATGATACACAACAACACCATTTCAA 134
Db 721 AAGGCCAGGTGTGACCATAGAAACCTGTATGATACACAACAACACCATTTCAA 780
Oy 135 AATGTTACCGACACACAGCCCAACACAGGATGTGTGACGCAATTTCTGAAC 194
Db 781 AATGTTACCGACACACAGCCCAACACAGGATGTGTGACGCAATTTCTGAAC 840
Oy 195 GTCCAGATATGTACTCTTGTGACACCTTTACCTTAAACATGCAAGAGATTTCA 254
Db 841 GTCCAGATATGTACTCTTGTGACACCTTTACCTTAAACATGCAAGAGATTTCA 900
Oy 255 CAACCTAAGAAACAGAAAGAAACAAAGAACCGGACCTTTGGAGCCGGAATGACAC 314
Db 901 CAACCTAAGAAACAGAAAGAAACAAAGAACCGGACCTTTGGAGCCGGAATGACAC 960
Oy 315 ACTGCCCCGAGGCTCAACCCCTGCTGTCTAAGAAATCTGGAATGTGGTCTCTTG 374
Db 961 ACTGCCCCGAGGCTCAACCCCTGCTGTCTAAGAAATCTGGAATGTGGTCTCTTG 1020
Oy 375 CTGCTTCAGCAGCTCTGTGATTAAACAGAGAGGATTCCTCCGAGAGAAGTGA 434
Db 1021 CTGCTTCAGCAGCTCTGTGATTAAACAGAGAGGATTCCTCCGAGAGAAGTGA 1080
Oy 435 GCTACGTTAGCCAGAGAAGAAAGGATGTGGTGAGAGGAAAGACTTCGACGACGTG 494
Db 1081 GCTACGTTAGCCAGAGAAGAAAGGATGTGGTGAGAGGAAAGACTTCGACGACGTG 1140
Oy 495 CAAGAATCAAGCCTTCATTAATTCCTGTCATCAT 532
Db 1141 CAAGAATCAAGCCTTCATTAATTCCTGTCATCAT 1178

```

```

RESULT 6
LOCUS BG974340 874 bp mRNA linear EST 12-JUN-2001
DEFINITION 602844094P1 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4979855 5',
RNA sequence.
ACCESSION BG974340
VERSION BG974340.1 GI:14361977
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 874)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furch
Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10978 row: k column: 24
High quality sequence stop: 846.
Location/Qualifiers
1. 874
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:4979855"
/clone_lib="NIH_CGAP_Mam4"
/tissue_type="tumor, gross tissue"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1; NCI;"

```

Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert 2.5 kb. Library constructed by Life  
Technologies, catalog # 12018-016. Investigators providing  
samples: Lohar Hennighausen/Priscilla Firth, NIH  
Reference for transgenic model: Li et al., Cell Growth and  
Differentiation 7, 3-11 (1996). Note: this is a NCI\_CGAP  
Library."

BASE COUNT 277 a 216 c 193 g 188 t  
ORIGIN

Query Match 96.8%; Score 514.8; DB 10; Length 874;  
Best Local Similarity 99.6%; Pred. No. 6.9e-105;  
Matches 516; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 15 ACACCCCTGCACAGAGATGTCAGTGTGCTTAGAGACAAGTCACTTGTCC 74  
DB 200 ACACCCCTGCACAGAGATGTCAGTGTGCTTAGAGACAAGTCACTTGTCC 259  
OY 75 AAGGCCAGGTGTACCATAGAAAACCTGTGTATGATCAACACACACATTCNA 134  
DB 260 AAGGCCAGGTGTACCATAGAAAACCTGTGTATGATCAACACACACATTCNA 319  
OY 135 AATGTTACCCGACACACAGCCCAACACAGGCGTGGGAAGCCATTTCGAC 194  
DB 320 AATGTTACCCGACACACAGCCCAACACAGGCGTGGGAAGCCATTTCGAC 379  
OY 195 GTCCAGATATGTACTCTTGTACCACTTACCTAAACATGCAAGAGATTTCTA 254  
DB 380 GTCCAGATATGTACTCTTGTACCACTTACCTAAACATGCAAGAGATTTCTA 439  
OY 255 CAACCTAAGAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 314  
DB 440 CAACCTAAGAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 499  
OY 315 ACTGCCCCAGGCGTCAACCCCTGCTTGTAAAGAAATCTGGAAATGTGGTCTTG 374  
DB 500 ACTGCCCCAGGCGTCAACCCCTGCTTGTAAAGAAATCTGGAAATGTGGTCTTG 559  
OY 375 CTGCTTACGAGAGTCTGTATTAACCAAGAGAGTTCCTCCGAGAGAAAGTGA 434  
DB 560 CTGCTTACGAGAGTCTGTATTAACCAAGAGAGTTCCTCCGAGAGAAAGTGA 619  
OY 435 GCTACGTAGCCAGAAAGAAAGAGATGGGTAGAGAGAAAGACTTGCAGGAGCTG 494  
DB 620 GCTACGTAGCCAGAAAGAAAGAGATGGGTAGAGAGAAAGACTTGCAGGAGCTG 679  
OY 495 CAAAGAAATCAAGCCTTTCATTAATTCCTGTCATCAT 532  
DB 680 CAAAGAAATCAAGCCTTTCATTAATTCCTGTCATCAT 717

RESULT 7 841 bp mRNA linear EST 29-MAY-2001  
LOCUS BG867642  
DEFINITION 602787429F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:4913382 5',  
ACCESSION BG867642  
VERSION BG867642.1 GI:14218182  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 841)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-r@mail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLNL0818 row: j column: 07  
High quality sequence stop: 801.  
Location/Qualifiers  
1. 841

FEATURES  
source

BASE COUNT 266 a 215 c 178 g 181 t  
ORIGIN

Query Match 96.8%; Score 513.8; DB 10; Length 841;  
Best Local Similarity 99.4%; Pred. No. 1.2e-104;  
Matches 515; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 15 ACACCCCTGCACAGAGATGTCAGTGTGCTTAGAGACAAGTCACTTGTCC 74  
DB 146 ACACCCCTGCACAGAGATGTCAGTGTGCTTAGAGACAAGTCACTTGTCC 205  
OY 75 AAGGCCAGGTGTACCATAGAAAACCTGTGTATGATCAACACACACATTCNA 134  
DB 206 AAGGCCAGGTGTACCATAGAAAACCTGTGTATGATCAACACACACATTCNA 265  
OY 135 AATGTTACCCGACACACAGCCCAACACAGGCGTGGGAAGCCATTTCGAC 194  
DB 266 AATGTTACCCGACACACAGCCCAACACAGGCGTGGGAAGCCATTTCGAC 325  
OY 195 GTCCAGATATGTACTCTTGTACCACTTACCTAAACATGCAAGAGATTTCTA 254  
DB 326 GTCCAGATATGTACTCTTGTACCACTTACCTAAACATGCAAGAGATTTCTA 385  
OY 255 CAACCTAAGAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 314  
DB 386 CAACCTAAGAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 445  
OY 315 ACTGCCCCAGGCGTCAACCCCTGCTTGTAAAGAAATCTGGAAATGTGGTCTTG 374  
DB 446 ACTGCCCCAGGCGTCAACCCCTGCTTGTAAAGAAATCTGGAAATGTGGTCTTG 505  
OY 375 CTGCTTACGAGAGTCTGTATTAACCAAGAGAGTTCCTCCGAGAGAAAGTGA 434  
DB 506 CTGCTTACGAGAGTCTGTATTAACCAAGAGAGTTCCTCCGAGAGAAAGTGA 565  
OY 435 GCTACGTAGCCAGAAAGAAAGAGATGGGTAGAGAGAAAGACTTGCAGGAGCTG 494  
DB 566 GCTACGTAGCCAGAAAGAAAGAGATGGGTAGAGAGAAAGACTTGCAGGAGCTG 625  
OY 495 CAAAGAAATCAAGCCTTTCATTAATTCCTGTCATCAT 532  
DB 626 CAAAGAAATCAAGCCTTTCATTAATTCCTGTCATCAT 663

RESULT 8 806 bp mRNA linear EST 21-NOV-2000  
LOCUS BF299645  
DEFINITION 602029327F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:4164463 5',  
ACCESSION BF299645  
VERSION BF299645.1 GI:11246168  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 806)  
 NIH-MGC <http://mgc.nci.nih.gov/>  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapds-remail.nih.gov  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLM9450 row: e column: 08  
 High quality sequence stop: 724.

FEATURES  
 source  
 Location/Qualifiers  
 1..806  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:416463"  
 /clone\_lib="NCI\_CGAP\_SG2"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: salivary gland; Vector: pCMV-Sport6; Site: 1; NotI: site 2; SalI: cloned unidirectionally. Primer: Oligo dT. Average insert size 1.3 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT 249 a 218 c 177 g 162 t  
 ORIGIN

Query Match 96.2%; Score 511.6; DB 10; Length 806;  
 Best Local Similarity 99.2%; Pred. No. 3.6e-104;  
 Matches 514; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

15 ACACCCCTGTCACAGAAATGTCAGTGTGCTTGTGAGACAAAGCACTTGTCC 74  
 100 ACACCCCTGTCACAGAAATGTCAGTGTGCTTGTGAGACAAAGCACTTGTCC 159  
 75 AAGGCCAGGTGTGACCATAGAAACTGTGTATGATCAACAACAACACCATTC 134  
 160 AAGGCCAGGTGTGACCATAGAAACTGTGTATGATCAACAACAACACCATTC 219  
 135 AATGTTACCGCAGACACAGCCCAACACAGGAGGTGTGGAAGCCAAATTCG 194  
 220 AATGTTACCGCAGACACAGCCCAACACAGGAGGTGTGGAAGCCAAATTCG 279  
 195 GTCCAGATATGATCTTCTGTACACCTTACTTAAATAAATGCAAGAGGATCT 254  
 280 GTCCAGATATGATCTTCTGTACACCTTACTTAAATAAATGCAAGAGGATCT 339  
 255 CAACCTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 314  
 340 CAACCTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 399  
 315 ACTGCCCCAGGGCTACCCCTGCTGTCTAAGAAATCTCGAATGTGGTCTCT 374  
 400 ACTGCCCCAGGGCTACCCCTGCTGTCTAAGAAATCTCGAATGTGGTCTCT 459  
 375 CTGCTTAAAGGAGTCTGTGATTAACCCAGAGGGAGTCTCTCCGAGAGAA 434  
 460 CTGCTTAAAGGAGTCTGTGATTAACCCAGAGGGAGTCTCTCCGAGAGAA 519  
 435 GCTACGTAGCCAGAGAAAGAGGATGTGGTGAAGAGGAAAGAAAGTTCG 494  
 520 GCTACGTAGCCAGAGAAAGAGGATGTGGTGAAGAGGAAAGAAAGTTCG 579  
 495 CAAGAAATCAAGCTTTCATTAATTCCTGTCATCAT 532  
 580 CACAGAAATCAAGCTTTCATTAATTCCTGTCATCAT 617

RESULT 9  
 BF300481

LOCUS BF300481 822 bp mRNA linear EST 21-NOV-2000  
 DEFINITION 602031613F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:416858 5',  
 mRNA sequence.  
 ACCESSION BF300481  
 VERSION BF300481.1 GI:11246991  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 822)  
 NIH-MGC <http://mgc.nci.nih.gov/>  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapds-remail.nih.gov  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLM9456 row: 1 column: 03  
 High quality sequence stop: 782.

FEATURES  
 source  
 Location/Qualifiers  
 1..822  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:416858"  
 /clone\_lib="NCI\_CGAP\_SG2"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: salivary gland; Vector: pCMV-Sport6; Site: 1; NotI: site 2; SalI: cloned unidirectionally. Primer: Oligo dT. Average insert size 1.3 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT 262 a 204 c 180 g 176 t  
 ORIGIN

Query Match 95.0%; Score 505.4; DB 10; Length 822;  
 Best Local Similarity 99.6%; Pred. No. 8.6e-103;  
 Matches 517; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

15 ACACCCCTGTCACAGAAATGTCAGTGTGCTTGTGAGACAAAGCACTTGTCC 74  
 119 ACACCCCTGTCACAGAAATGTCAGTGTGCTTGTGAGACAAAGCACTTGTCC 178  
 75 AA-GGCCAGGTGTGACCATAGAAACTGTGTATGATCAACAACAACACCATTC 133  
 179 AAGGCCAGGTGTGACCATAGAAACTGTGTATGATCAACAACAACACCATTC 238  
 134 AATGTTACCGCAGACACAGCCCAACACAGGAGGTGTGGAAGCCAAATTCG 193  
 239 AATGTTACCGCAGACACAGCCCAACACAGGAGGTGTGGAAGCCAAATTCG 298  
 194 GTCCAGATATGATCTTCTGTACACCTTACTTAAATAAATGCAAGAGGATCT 253  
 299 GTCCAGATATGATCTTCTGTACACCTTACTTAAATAAATGCAAGAGGATCT 358  
 254 ACAACCTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 313  
 359 ACAACCTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 418  
 314 CACTGCCCCAGGGCTACCCCTGCTGTCTAAGAAATCTCGAATGTGGTCTCT 373  
 419 CACTGCCCCAGGGCTACCCCTGCTGTCTAAGAAATCTCGAATGTGGTCTCT 478  
 374 GTGCTTAAAGGAGTCTGTGATTAACCCAGAGGGAGTCTCTCCGAGAGAA 433  
 479 GTGCTTAAAGGAGTCTGTGATTAACCCAGAGGGAGTCTCTCCGAGAGAA 538  
 434 AGCTACGTAGCCAGAGAAAGAGGATGTGGTGAAGAGGAAAGAAAGTTCG 493

```

|||||
Db 539 ACCTAGCTTAGCCAGAGAGAAAGGATGGTGAGAGAGAAAGACTTCGACGACGT 598
OY 494 GCAAAATCAAGCTTCATTAATTCCTGTCAATCAT 532
Db 599 GCAAAAGAAATCAAGCTTCATTAATTCCTGTCAATCAT 637

RESULT 10
Bg666897 740 bp mRNA linear EST 29-MAY-2001
LOCUS 602786139P1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4912329 5',
DEFINITION mRNA sequence.
ACCESSION Bg666897
VERSION Bg666897.1 GI:14217437
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 740)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabds-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM10815 row: n column: 10
High quality sequence stop: 728.
Location/Qualifiers
1..740
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4912329"
/lab_host="NCI_CGAP_SG2"
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site: 1;
NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 230 a 189 c 158 g 163 t
ORIGIN
Query Match 94.8%; Score 504.4; DB 10; Length 740;
Best Local Similarity 99.6%; Pred. No. 1.4e-102;
Matches 516; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 15 ACACCCCCCGACAGAGATGTCAGTGGTGTGTTGAGACAAAGTCACCTTTGTC 74
Db 46 ACACCCCCCGACAGAGATGTCAGTGGTGTGTTGAGACAAAGTCACCTTTGTC 105
OY 75 AAGGCCCAGGTGTGACCATAGAAACTGTATTGATGATCAACAACACACCATTCAA 134
Db 106 AAGGCCCAGGTGTGACCATAGAAACTGTATTGATGATCAACAACACACCATTCAA 165
OY 135 AATGGTTACCGACACACAGCCCAACACACAGGTTGTAAGCCAAATTCCTGAAAC 194
Db 166 AATGGTTACCGACACACAGCCCAACACACAGGTTGTAAGCCAAATTCCTGAAAC 225
OY 195 GTCCAGATATGTACTTCTGTACCACTTACCTAAAAAATGCGAGAGAGATTCCTA 254
Db 226 GTCCAGATATGTACTTCTGTACCACTTACCTAAAAAATGCGAGAGAGATTCCTA 285
OY 255 CAACCTTAAGAAAGAAAGAAAGAAAGAACCGGACCTTTTGAAGCGGAAATGACAC 314
Db 286 CAACCTTAAGAAAGAAAGAAAGAAAGAACCGGACCTTTTGAAGCGGAAATGACAC 345

```

```

OY 315 ACTGGCCCCAGGCTCACCCCTCCCTTGTCTTAAGAAATCTGGATGTGGTCTCTTG 374
Db 346 ACTGGCCCCA-GGCTCACCCCTCCCTTGTCTTAAGAAATCTGGATGTGGTCTCTTG 404
OY 375 CTGGCTTCAGGATCTCTGTATTACCCAGAGGAGTTCCTCCGAGAGAGAAAGTGA 434
Db 405 CTGGCTTCAGGATCTCTGTATTACCCAGAGGAGTTCCTCCGAGAGAAAGTGA 464
OY 435 GCTAGCTTAGCCAGAGAGAAAGGATGTGGTGAGAGGAGAAAGCTTCAGAGACGTG 494
Db 465 GCTAGCTTAGCCAGAGAGAGAAAGGATGTGGTGAGAGGAGAAAGCTTCAGAGACGTG 524
OY 495 CAAGAAGATCAAGCTTCATTAATTCCTGTCAATCAT 532
Db 525 CAAGAAGATCAAGCTTCATTAATTCCTGTCAATCAT 562

RESULT 11
BF537367 849 bp mRNA linear EST 11-DEC-2000
LOCUS 602050004F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4189464 5',
DEFINITION mRNA sequence.
ACCESSION BF537367
VERSION BF537367.1 GI:11624735
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 849)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabds-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM9515 row: g column: 01
High quality sequence stop: 828.
Location/Qualifiers
1..849
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4189464"
/lab_host="NCI_CGAP_SG2"
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site: 1;
NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 271 a 207 c 193 g 178 t
ORIGIN
Query Match 94.8%; Score 504.4; DB 10; Length 849;
Best Local Similarity 99.6%; Pred. No. 1.4e-102;
Matches 516; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 15 ACACCCCCCGACAGAGATGTCAGTGGTGTGTTGAGACAAAGTCACCTTTGTC 74
Db 135 ACACCCCCCGACAGAGATGTCAGTGGTGTGTTGAGACAAAGTCACCTTTGTC 193
OY 75 AAGGCCCAGGTGTGACCATAGAAACTGTATTGATGATCAACAACACACCATTCAA 134
Db 194 AAGGCCCAGGTGTGACCATAGAAACTGTATTGATGATCAACAACACACCATTCAA 253
OY 135 AATGGTTACCGACACACAGCCCAACACACAGGTTGTAAGCCAAATTCCTGAAAC 194

```

```

Db 254 AATGTTACCGCAGACACAGCCCAACACAGGAGTGTGAGACCAATTTCTGAAAC 313
Oy 195 GTCCAGATATGATCTTGTGTACCACTTTACCTAAAAAATGAGAGAGATTTCTA 254
Db 314 GTCCGATATGATCTTGTGTACCACTTTACCTAAAAAATGAGAGAGATTTCTA 373
Oy 255 CAACCTAAGAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 314
Db 374 CAACCTAAGAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 433
Oy 315 ACTGGCCCCAGGGCTCACCTGCCCCCTTGTCTAAGAAATCTGGAATGTGGCTCTCTG 374
Db 434 ACTGGCCCCAGGGCTCACCTGCCCCCTTGTCTAAGAAATCTGGAATGTGGCTCTCTG 493
Oy 375 CTGCTTCAGCGAGTCTGTATTAACCAAGAGAGAGTTCCTCCGAGAGAAAGTGGGA 434
Db 494 CTGCTTCAGCGAGTCTGTATTAACCAAGAGAGAGTTCCTCCGAGAGAAAGTGGGA 553
Oy 435 GCTACGTTAGCCAGAAAGAGAGAGATGGGTGAGAGAGAAAGAAAGAAAGTGGGA 494
Db 554 GCTACGTTAGCCAGAAAGAGAGAGATGGGTGAGAGAGAAAGAAAGTGGGA 613
Oy 495 CAAGAATATCAAGCCTTTCAATATTTCCCTGTCATCAT 532
Db 614 CAAGAATATCAAGCCTTTCAATATTTCCCTGTCATCAT 651

```

RESULT 12  
BF535022 891 bp mRNA linear EST 11-DEC-2000  
LOCUS 602050476F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:418986 5'

DEFINITION mRNA sequence.  
ACCESSION BF535022  
VERSION BF535022.1 GI:11622385  
KEYWORDS EST.

SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 891)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov

Plate: LLM9516 row: g column: 19  
High quality sequence stop: 739.  
Location/Qualifiers  
1. 891  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone\_image="418986"  
/clone\_lib="NCI\_CGAP\_SG2"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: salivary gland; Vector: pCMV-SPORT6, site\_1: NotI; site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.3 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP library."

## FEATURES

source

BASE COUNT 273 a 218 c 223 g 177 t  
ORIGIN

Query Match 94.8%; Score 504.4; DB 10; Length 891;  
Best Local Similarity 99.6%; Pred. No. 1.4e-102;  
Matches 516; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```

Oy 15 ACACCCCTGACAGAGATGTCAGTGTGTGTGTTAGAGCAAAAGTCATTTGTC 74
Db 94 ACACCCCTGACAGAGATGTCAGTGTGTGTGTTAGAGCAAAAGTCATTTGTC 153
Oy 75 AAGGCCAGAGTGTGTGACCATGAAAACTGTGTATGATTCACAAACACACCATTCAA 134
Db 154 AAGGCCAGAGTGTGTGACCATGAAAACTGTGTATGATTCACAAACACACCATTCAA 213
Oy 135 AATGGTTACCGCAGACACAGCCCAACACAGAGGTGTGAGAGCAATTTCTGAAAC 194
Db 214 AATGGTTACCGCAGACACAGCCCAACACAGAGGTGTGAGAGCAATTTCTGAAAC 273
Oy 195 GTCCAGATATGATCTTGTGTACCACTTTACCTAAAAAATGAGAGAGATTTCTA 254
Db 274 GTCCGATATGATCTTGTGTACCACTTTACCTAAAAAATGAGAGAGATTTCTA 333
Oy 255 CAACCTAAGAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 314
Db 334 CAACCTAAGAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 392
Oy 315 ACTGGCCCCAGGGCTCACCTGCCCCCTTGTCTAAGAAATCTGGAATGTGGCTCTCTG 374
Db 393 ACTGGCCCCAGGGCTCACCTGCCCCCTTGTCTAAGAAATCTGGAATGTGGCTCTCTG 452
Oy 375 CTGCTTCAGCGAGTCTGTATTAACCAAGAGAGAGTTCCTCCGAGAGAAAGTGGGA 434
Db 453 CTGCTTCAGCGAGTCTGTATTAACCAAGAGAGAGTTCCTCCGAGAGAAAGTGGGA 512
Oy 435 GCTACGTTAGCCAGAAAGAGAGATGGGTGAGAGAGAAAGAAAGTGGGA 494
Db 513 GCTACGTTAGCCAGAAAGAGAGATGGGTGAGAGAGAAAGAAAGTGGGA 572
Oy 495 CAAGAATATCAAGCCTTTCAATATTTCCCTGTCATCAT 532
Db 573 CAAGAATATCAAGCCTTTCAATATTTCCCTGTCATCAT 610

```

RESULT 13  
BG873314 743 bp mRNA linear EST 29-MAY-2001  
LOCUS 602794169F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:4925297 5'

DEFINITION mRNA sequence.  
ACCESSION BG873314  
VERSION BG873314.1 GI:14223854  
KEYWORDS EST.

SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 743)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LLM10849 row: j column: 18  
High quality sequence stop: 681.  
Location/Qualifiers  
1. 743  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone\_image="4925297"  
/clone\_lib="NCI\_CGAP\_SG2"  
/lab\_host="DH10B (T1 phage-resistant)"

## FEATURES

source



/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site: 1; NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.3 kb. Constructed by Life Technologies. Note: this is a NCI-CGAP Library."

BASE COUNT 226 a 187 c 172 g 158 t

ORIGIN

Query Match 94.7%; Score 503.8; DB 10; Length 743;  
Best Local Similarity 99.4%; Pred. No. 2e-102;  
Matches 516; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

15 ACACCCCTGCACAGAGATGTCAGTGGTTGCTTACAGCAAGTCACTTGTCC 74  
|||||  
31 ACACCCCTGCACAGAGATGTCAGTGGTTGCTTACAGCAAGTCACTTGTCC 90  
75 AAGGCCAGGTGTGACCATATGAAACCTGTTATGATCAACACACACCATTC 134  
91 AAGGCCAGGTGTGACCATATGAAACCTGTTATGATCAACACACACCATTC 150  
135 AATGTTACCGCAGACACACAGCCCAACACAGGGTGTGGAAGCCAAATTCCT 194  
151 AATGTTACCGCAGACACACAGCCCAACACAGGGTGTGGAAGCCAAATTCCT 210  
195 GTCCAGATATGATCTCTGTTACACCTTACTTAAACATGCAAGAGATTCCT 254  
211 GTCCAGATATGATCTCTGTTACACCTTACTTAAACATGCAAGAGATTCCT 270  
255 CAACCTAAGAAACAGAGAAACAAAGAAACCGGCACTTTGGAGCCGAATTC 314  
271 CAACCTAAGAAACAGAGAAACAAAGAAACCGGCACTTTGGAGCCGAATTC 330  
315 ACTGAGCCCGAGGCTCACTCCCTGTTCTAAGAAATCCTGGAATGTGGTCT 374  
331 ACTGAGCCCGAGGCTCACTCCCTGTTCTAAGAAATCCTGGAATGTGGTCT 390  
375 CTGCTTACAGGAGTCTGATTAACCCAGAGAGTCTCTCCGAGAGAAATGGA 434  
391 CTGCTTACAGGAGTCTGATTAACCCAGAGAGTCTCTCCGAGAGAAATGGA 450  
435 GCTACGTTAGCCAGAGAGAA-AGGATGTGGTGAAGAGAAAGACTTCGAGG 493  
451 GCTACGTTAGCCAGAGAGAAAGGAGATGTGGTGAAGAGAAAGACTTCGAGG 510  
494 GCAAGAAATCAAGCCTTTCATTAATTCCTGCAATCAT 532  
511 GCAAGAAATCAAGCCTTTCATTAATTCCTGCAATCAT 549

RESULT 14  
BF300873 988 bp mRNA linear EST 21-NOV-2000  
LOCUS 602028873F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:416394 5',  
DEFINITION mRNA sequence.  
ACCESSION BF300873  
VERSION BF300873.1 GI:11247396  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 988)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/MLNI at:  
http://image.llnl.gov

Plate: LLAM9449 row: a column: 19  
High quality sequence stop: 743.  
Location/Qualifiers  
1. '988  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:416394"  
/clone\_11b="NCI\_CGAP\_SG2"  
/lab\_host="DH10B (TI phage-resistant)"  
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site: 1; NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.3 kb. Constructed by Life Technologies. Note: this is a NCI-CGAP Library."

BASE COUNT

319 a 244 c 241 g 184 t

ORIGIN

Query Match 94.5%; Score 502.8; DB 10; Length 988;  
Best Local Similarity 99.4%; Pred. No. 3.3e-102;  
Matches 515; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

15 ACACCCCTGCACAGAGATGTCAGTGGTTGCTTACAGCAAGTCACTTGTCC 74  
|||||  
119 ACACCCCTGCACAGAGATGTCAGTGGTTGCTTACAGCAAGTCACTTGTCC 178  
75 AAGGCCAGGTGTGACCATATGAAACCTGTTATGATCAACACACACCATTC 134  
179 AAGGCCAGGTGTGACCATATGAAACCTGTTATGATCAACACACACCATTC 150  
213 AATGTTACCGCAGACACACAGCCCAACACAGGGTGTGGAAGCCAAATTCCT 194  
239 AATGTTACCGCAGACACACAGCCCAACACAGGGTGTGGAAGCCAAATTCCT 258  
195 GTCCAGATATGATCTCTGTTACACCTTACTTAAACATGCAAGAGATTCCT 254  
211 GTCCAGATATGATCTCTGTTACACCTTACTTAAACATGCAAGAGATTCCT 270  
255 CAACCTAAGAAACAGAGAAACAAAGAAACCGGCACTTTGGAGCCGAATTC 314  
271 CAACCTAAGAAACAGAGAAACAAAGAAACCGGCACTTTGGAGCCGAATTC 330  
315 ACTGAGCCCGAGGCTCACTCCCTGTTCTAAGAAATCCTGGAATGTGGTCT 374  
331 ACTGAGCCCGAGGCTCACTCCCTGTTCTAAGAAATCCTGGAATGTGGTCT 390  
375 CTGCTTACAGGAGTCTGATTAACCCAGAGAGTCTCTCCGAGAGAAATGGA 434  
391 CTGCTTACAGGAGTCTGATTAACCCAGAGAGTCTCTCCGAGAGAAATGGA 450  
435 GCTACGTTAGCCAGAGAGAAAGGAGATGTGGTGAAGAGAAAGACTTCGAGG 494  
451 GCTACGTTAGCCAGAGAGAAAGGAGATGTGGTGAAGAGAAAGACTTCGAGG 510  
494 GCAAGAAATCAAGCCTTTCATTAATTCCTGCAATCAT 532  
511 GCAAGAAATCAAGCCTTTCATTAATTCCTGCAATCAT 549

RESULT 15  
BG871007 741 bp mRNA linear EST 29-MAY-2001  
LOCUS 602792051F1 NCI\_CGAP\_SG2 Mus musculus cDNA clone IMAGE:4923253 5',  
DEFINITION mRNA sequence.  
ACCESSION BG871007  
VERSION BG871007.1 GI:14221547  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 741)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL  
COMMENT

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-r@mail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINT)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINT at:  
<http://image.llnl.gov>  
Plate: LLM10844 row: e column: 14  
High quality sequence stop: 740.  
Location/Qualifiers  
1. 741

FEATURES  
source

/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4923253"  
/lab\_host="NCL-CGAP-S62"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: Salivary gland; Vector: PCMV-SPORT6; Site\_1:  
Not1; Site\_2: Sal1; Cloned unidirectionally. Primer: Oligo  
dT. Average insert size 1.3 kb. Constructed by Life  
Technologies. Note: this is a NCL-CGAP Library."  
BASE COUNT 231 a 192 c 166 g 152 t  
ORIGIN

Query Match 93.3%; Score 496.4; DB 10; Length 741;  
Best local similarity 98.6%; Pred. No. 8.8e-101;

Matches 511; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 15 ACACCCCTCCGACAGAGATGCCAGTGTGTGCTTAGAGACAAAGTCATTGTGC 74  
DB 201 ACACCCCTCCGACAGAGATGCCAGTGTGTGCTTAGAGACAAAGTCATTGTGC 260  
QY 75 AAGGCCAGGTGTGATGAAACCTGTATGATCACAACAAACACCATTCAC 134  
DB 261 AAGGCCAGGTGTGATGAAACCTGTATGATCACAACAAACACCATTCAC 320  
QY 135 AATGTTACCGACACACAGCCCAACACAGGAGTGAGAAAGCCATTTCTGAAC 194  
DB 321 AATGTTACCGACACACAGCCCAACACAGGAGTGAGAAAGCCATTTCTGAAC 380  
QY 195 GTCCAGATATGTACTTCTGTACACCTTACTTAAAAACATGACAGAGATTCTA 254  
DB 381 GTCCAGATATGTACTTCTGTACACCTTACTTAAAAACATGACAGAGATTCTA 440  
QY 255 CAACCTAAGAAACAGAAACAAAGAAACGCGCACTTTGGAGCGGAATGACAC 314  
DB 441 CAACCTAAGAAACAGAAACAAAGAAACGCGCACTTTGGAGCGGAATGACAC 500  
QY 315 ACTGCCCCAGGGCTCACCTGCCCCCTGTAAAGAACTCTGAATGTGGTCTCTTG 374  
DB 501 ACTGCCCCCA-GGCTCACCTGCCCCCTGTAAAGAACTCTGAATGTGGTCTCTTG 559  
QY 375 CTGCTTACAGGAGTCTGATTAACCAAGAGAGTCTCTCCGAGAGAAAGTGA 434  
DB 560 CTGCTTACAGGAGTCTGATTAACCAAGAGAGTCTCTCCGAGAGAAAGTGA 619  
QY 435 GGTAGTTAGCCAGAAAGAAAGGATGGGTGAGAGAAAGACTTCAGCGAGCTG 494  
DB 620 GGTAGTTAGCCAGAAAGAAAGGATGGGTGAGAGAAAGACTTCAGCGAGCTG 679  
QY 495 CAAGAAATCAAGCTTTCATTAATCCCTGCAATCAT 532  
DB 680 CAAGAAATCAAGCTTTCATTAATCCCTGCAATCAT 717

Search completed: October 19, 2002, 06:53:28  
Job time : 1643 secs